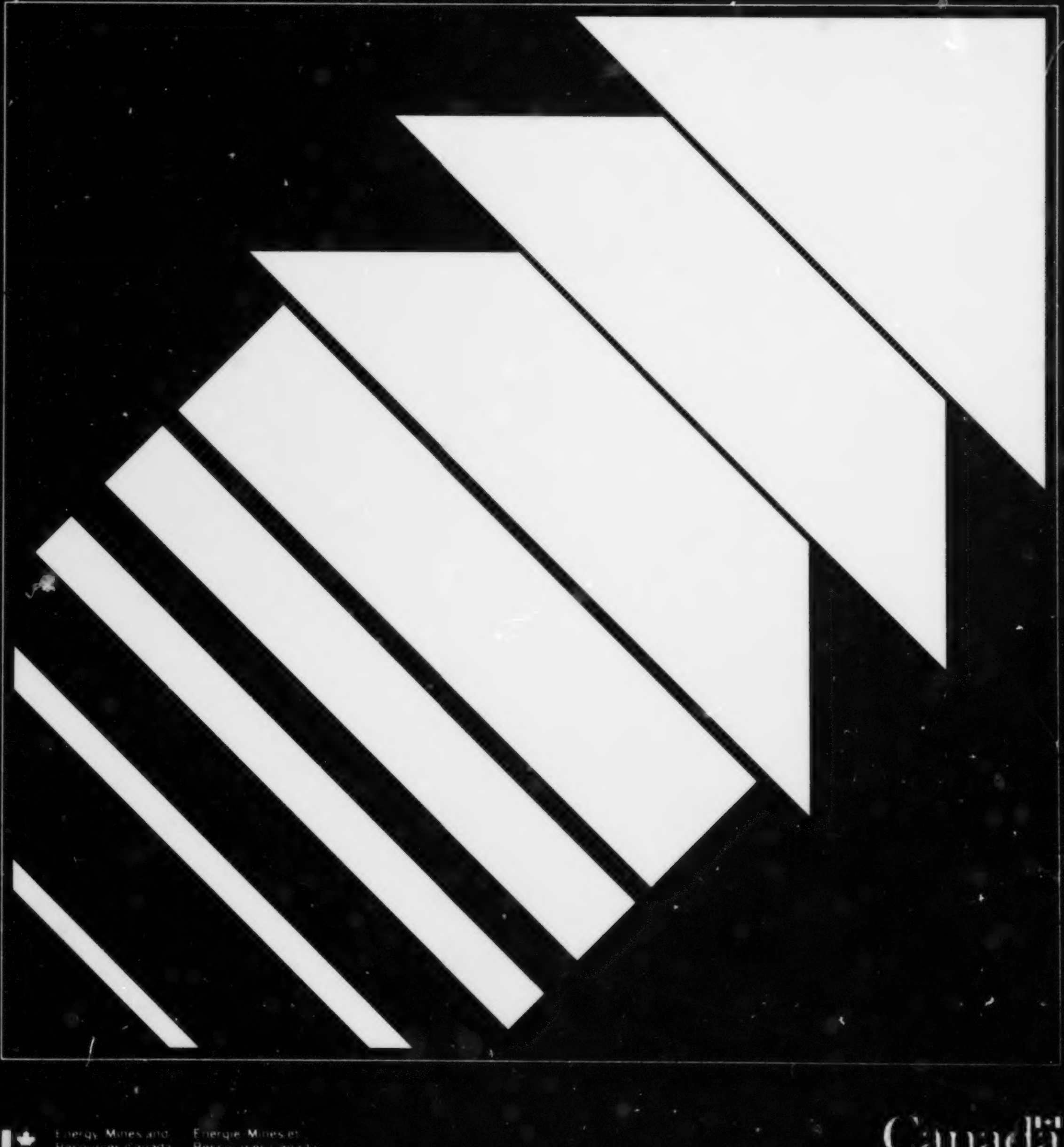
NATIONAL URANIUM TAILINGS PROGRAM



Energy, Mines and Resources Canada
Énergie, Mines et Ressources Canada

Canada

PROBABILISTIC MODEL DEVELOPMENT FOR THE ASSESSMENT OF THE LONG-TERM EFFECTS OF URANIUM MILL TAILINGS IN CANADA - PHASE II

DSS File #26SQ.23241-4-1693
Contract #0SQ84-00207

VOLUME 1 - MAIN REPORT

A Research Report Prepared for:

The National Uranium Tailings Program
Canada Centre for Mineral and Energy Technology
Energy, Mines and Resources Canada
555 Booth Street
Ottawa, Ontario
K1A 0G1

by

SENES Consultants Limited
499 McNicoll Avenue
Willowdale, Ontario
M2H 2C9

October 1985

153 pp

MICROMEDIA

ABSTRACT

This study was conducted to extend and refine the probabilistic assessment tool, developed in a previous study for assessing the long-term effects of uranium mill tailings disposal. The incorporation of time-dependent behaviour was a particularly important aspect of the study. Numerous technical issues relevant to performing or interpreting a probabilistic assessment were examined. A computer code was developed and used to perform probabilistic analyses on several uranium tailings management scenarios.

NOTICE

This report was prepared by SENES Consultants Limited (SENES) under contract to the National Uranium Tailings Program (NUTP). Neither the NUTP nor SENES makes any warranty, expressed or implied, or assumes any liability with respect to any damage or loss incurred as a result of the use made of the information contained in this publication.

TABLE OF CONTENTS

Page

ABSTRACT i
NOTICE i
EXECUTIVE SUMMARY 1

1.0 INTRODUCTION 1-1
  1.1 Scope of Phase II Investigations 1-2
  1.2 Report Structure 1-3

2.0 BACKGROUND CONSIDERATIONS 2-1
  2.1 Reference Tailings Site Characteristics 2-3
  2.2 The Reference Receptor 2-4
  2.3 Environmental Models 2-5
    2.3.1 The Tailings Source Term 2-6
    2.3.2 Environmental Dispersion 2-10
    2.3.3 Environmental Transfer Models 2-12
  2.4 Model Aggregation Studies 2-12
  2.5 Disposal Technologies Considered 2-13
  2.6 Some Aspects of Time Dependency 2-16
  2.7 Sensitivity and Uncertainty Analysis Techniques 2-19
    2.7.1 Sensitivity Index 2-21
    2.7.2 Stepwise Regression 2-22

3.0 SYSTEM CODE STRUCTURE 3-1
  3.1 Overview of the Systems Approach 3-1
  3.2 The Random Simulation Control Module 3-1
  3.3 Component Modules and Parameter Allocation and Control 3-4

Tables of Contents (continued)

3.4 The Quality Assurance Program 3-5
    3.4.1 Tests of the Random Simulation Tool RANSIM 3-6
    3.4.2 Component Model QA 3-9
    3.4.3 System Model QA 3-10

4.0 PHASE II PROBABILISTIC ASSESSMENT RESULTS 4-1
  4.1 Assessment Methodology 4-1
    4.1.1 Number of Trials 4-1
    4.1.2 Implementation of Time Dependency 4-2

  4.2 Scenario 1: The Base Case 4-2
    4.2.1 Overall Evalaution 4-3
    4.2.2 Further Considerations of Temporal and Spatial Dependence 4-6
    4.2.3 Importance Analysis: Total Dose 4-8
    4.2.4 Importance Analysis: Selected Pathways 4-10
    4.2.5 Scenario 1 Summary 4-11
  4.3 Scenario 2: Pyrite Reduction 4-11
    4.3.1 Overall Evaluation 4-12
    4.3.2 Scenario 2 Summary 4-14
  4.4 Scenario 3: Pyrite Reduction and Limestone Addition 4-14
    4.4.1 Overall Evaluation 4-15
  4.5 Scenario 4: Bulk Float and Limestone Addition 4-16
    4.5.1 Overall Evaluation 4-17

Tables of Contents (continued)

5.0 SUMMARY OBSERVATIONS FROM TOPICAL DISCUSSION PAPERS 5-1
  5.1 Some Aspects of Parameter Correlation 5-3
  5.2 Model Aggregation 5-7
  5.3 Sensitivity/Uncertainty Analysis Techniques 5-9
  5.4 Statistical Confidence in Monte Carlo Calculations 5-12

6.0 SUMMARY AND RECOMMENDATIONS 6-1
  6.1 Summary 6-1
  6.2 Recommendations for Further Work 6-3

REFERENCES R-1

Appendix A Detailed Results of Sensitivity Index Analysis and Stepwise Regression

Appendix B Detailed Listing of Input Parameters

LIST OF FIGURES

Figure		Follows Page
2.1	NUTP Approach to Model Development	2-1
2.2	Concept of Probabilistic Analysis	2-2
2.3	Schematic Illustration of Setting for Reference Tailings Site	2-3
2.4	Pathways Studies in Phase II	2-6
2.5	Concept Adopted for Modelling the Tailings Source Term	2-7
2.6	Conceptual Approach to Sampling Time-Varying Parameters	2-19
3.1	Conceptual Flow Diagram of Systems Approach	3-1
3.2	Conceptual Flow Diagram of RANSIM	3-2
3.3	Conceptual Flow Diagram for PAC	3-4
3.4	Plot of Frequency Histogram (500 Samples) and Probability Distribution (Smooth Curve) for Normal Distribution.	3-8
3.5	Plot of Frequency Histogram (500 Samples) and Probability Distribution (Smooth Curve) for Lognormal Distribution	3-8
3.6	Plot of Frequency Histogram (500 Samples) and Probability Distribution (Smooth Curve) for Uniform Distribution	3-8
3.7	Plot of Frequency Histogram (500 Samples) and Probability Distribution (Smooth Curve) for Triangular Distribution	3-8
4.1	Conceptual Structure of Component Model Integration	4-1
4.2	Dose vs. Time for Scenario 1: Geometric Mean of 500 Trials	4-3

List of Figures (continued)

Figure		Follows Page
4.3	pH vs. Time for Scenario 1	4-5
4.4a)	Histogram of Total Dose at T=0 Years for Scenario 1	4-5
4.4b)	Histogram of Total Dose at T=200 Years for Scenario 1	4-5
4.5	Histogram of Dose from Radon Inhalation at T=0 Years for Scenario 1	4-5
4.6	Histogram of Dose from Drinking Water at T=200 Years for Scenario 1	4-5
4.7	Temporal Considerations in Probabilistic Analysis	4-6
4.8	Spatial Considerations in Probabilistic Analysis	4-7
4.9	Change in Parameter Importance Ranking vs Time	4-9
4.10	Contribution of Pathways to Total Dose and Identification of Important Parameters at T=0 Years	4-11
4.11	Contribution of Pathways to Total Dose and Identification of Important Parameters at T=200 Years	4-11
4.12	Dose vs. Time for Scenario 2: Geometric Mean of 500 Trials	4-12
4.13	Histogram of Total Dose at T=100 Years for Scenario 2	4-12
4.14	pH vs. Time for Scenario 2	4-14
4.15	Contribution of Pathways to Total Dose and Identification of Important Parameters at T=100 Years	4-14

List of Figures (continued)

Figure — Follows Page

4.16 Dose vs. Time for Scenario 3: Geometric Mean of 500 Trials 4-15
4.17a Histogram of Total Dose at T=0 Years for Scenario 3 4-15
4.17b Histogram of Total Dose at T=100 Years for Scenario 3 4-15
4.18 pH vs. Time for Scenario 3 4-16
4.19 Dose vs. Time for Scenario 4: Goemetric Mean of 500 Trials 4-17
4.20a Histogram of Total Dose at T=0 Years for Scenario 4 4-17
4.20b Histogram of Total Dose at T=50 Years for Scenario 4 4-17
4.21 pH vs. Time for Scenario 4 4-17
4.22 Total Dose vs. Time Summary; Geometric Mean of 500 Trials 4-18
4.23 pH vs. Time Summary 4-18

LIST OF TABLES

Table		Follows Page
2.1	Chemical and Radiological Characteristics of Uranium Tailings Surface Zone (Upper 1 m)	2-15
4.1	Parameter Importance Relative to Total Dose at T=0 Years for Scenario 1	4-8
4.2	Parameter Importance Relative to Total Dose at T=200 Years for Scenario 1	4-8
4.3	Ten Least-Important Parameters for Scenario 1	4-10
4.4	Parameter Importance Relative to Dose Via Water Consumption Pathway at T=200 Years for Scenario 1	4-10
4.5	Parameter Importance Relative to Dose Via Radon Inhalation at T=0 Years for Scenario 1	4-10
4.6	Parameter Importance Relative to pH at T=0 Years for Scenario 1	4-11
4.7	Parameter Importance Relative to Dose at T=100 Years for Scenario 2	4-13

EXECUTIVE SUMMARY

This study, carried out as part of the National Uranium Tailings Program, was conducted to extend and refine the probabilistic tools developed in the Phase I study and to investigate various issues of technical interest to performing and interpreting a probabilistic analysis.

In general terms the objectives of the Phase II investigations were to:

- incorporate time-dependent behaviour,
- include additional radionuclides,
- include non-radioactive constituents,
- define parameter distributions for transfer factors,
- evaluate conventional disposal technologies,
- carry out a preliminary assessment of a generic site,
- demonstrate the utility of sensitivity/uncertainty analysis techniques, and
- continue development of the system code.

In addition, several important technical issues were investigated including:

- interpretation of probabilistic results,
- methods for estimating confidence regions on output distributions,
- techniques for handling correlated parameters,
- model response to time-variant input parameters, and
- implications of model aggregation.

The incorporation of time-dependent behaviour was a particularly important aspect of the Phase II investigation. To this end, it was necessary to develop time-dependent component models and to restructure the system code. As expected, the time-dependent behaviour of the system was driven by physical and chemical changes in the tailings source term.

The source term, namely pyritic-uranium tailings, was modelled as a layered system; consisting of an upper, unsaturated zone and a lower, saturated zone. The oxidation of pyrite in the unsaturated zone, and the dissolution and reprecipitation of a gypsum-radium solid solution governed the behaviour of numerous other chemical species. Species modelled include radium, thorium, radon, calcium, iron and sulphate as well as pH.

Time considerations also gave rise to complications regarding the treatment of parameters which are not only uncertain, but vary from year to year. Mean annual windspeed is an example of a time-variant parameter. Several methods for dealing with such parameters were examined and are discussed in this report. The method employed during Phase II involved randomly sampling such parameters only once for each trial, from a distribution adjusted for the length of the simulation period.

The time-dependent component models, and revised control modules, were integrated to form the Uranium Tailings Assessment Program (UTAP). Documentation of this code and a user's manual for this program form part of this report.

The computer code (the system model) was used to perform probabilistic analyses of a number of different tailings management scenarios. The results of all analyses are summarized in this report.

Several sensitivity and uncertainty analyses techniques were examined to determine their applicability in distinguishing the contributions (importance) of the model parameters to the output predictions. Sensitivity analysis is used to assess how the model output changes with respect to changes in one or more input parameters. Uncertainty analysis is used to assess how much the uncertainty associated wtih each input parameter contributes to the uncertainty (spread) in the model predictions. The most applicable techniques were applied to the Phase II simulation results for the base case scenario (pyrite content of 6.5% in the tailings) and for the reduced pyrite disposal scenario (pyrite content of 0.6% in uppermost metre of the tailings).

In summary, all of the Phase II objectives have been fulfilled, all investigations were successfully carried out and the results are discussed in the appropriate sections of this report. The report is presented in three volumes, as indicated below, with many of the more technical discussions and details reserved for volumes 2 and 3.

Volume 1:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 1, Main Report.
Volume 2:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 2, Workshop Notes and Topical Discussion Papers.
Volume 3:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 3, Probabilistic Analysis Users Manual and Documentation for the Uranium Tailings Assessment Program (UTAP).

1.0 INTRODUCTION

In 1982 Energy, Mines and Resources Canada established the National Uranium Tailings Program (NUTP) which is intended to pull together previous work on Canadian uranium mill tailings, to fund additional work where significant information gaps are found and to develop the capability to predict the long-term effects of tailings. Such evaluations require the use of mathematical models to simulate the behaviour of the real system. In this regard, SENES Consultants Limited carried out a Phase I study for the NUTP to demonstrate how probabilistic techniques could be used to take account of the natural variability and uncertainty in model parameters in the simulation analysis (SENES 1984).

Subsequently, SENES undertook a Phase II study to extend and refine the probabilistic tools developed in the Phase I study and to investigate various issues of technical interest involved in carrying out and interpreting a probabilistic analysis. The findings of the Phase II SENES study are presented in a three-volume series of reports. This report, Volume 1 of the three volume set, is the main report and documents all of the principal findings of the Phase II studies. The contents of Volumes 2 and 3 are discussed in Section 1.2.

The first phase of work on modelling the overall system (SENES 1984) focussed on evaluating the applicability of probabilistic assessment techniques to modelling the long-term effects of uranium mill tailings for a simplified reference site. The results of this preliminary evaluation demonstrated the applicability of the probabilistic technique for estimating the magnitude and uncertainty of potential effects, such as annual dose to a specified nearby receptor.

During the Phase I investigation a systems level code with probabilistic analysis capabilities was selected and modified to permit application to a reference uranium tailings site. The analysis considered a limited set of site conditions and exposure pathways. In particular the following characteristics were incorporated:

- preliminary component and transfer models were developed for estimating human exposure from the release of radium-226 to the air and water environments and radon-222 to the air environment;
- standard tailings disposal technology was assumed involving tailings placement in a valley depression behind a relatively impervious dam;
- tailings were also assumed to be pyritic in nature and to remain exposed to natural weathering and oxidation processes.

Preliminary estimates of the respective component and transfer model parameter values and their associated statistical distributions were also developed. Using these distributions as the input specifications, a demonstration of the probabilistic assessment technique was carried out on the reference site. A preliminary sensitivity analysis was also carried out. In addition a number of topical papers were prepared on selected issues considered important to carrying out and evaluating a probabilistic assessment.

1.1 Scope of Phase II Investigations

The object of the Phase II studies is to build upon the results of the preliminary probabilistic assessment carried out in Phase I. In general terms the objectives of the Phase II studies were to:

- incorporate time-dependent behaviour,
- include additional radionuclides,
- include non-radioactive constituents,
- define transfer factor parameter distributions,
- evaluate conventional disposal technologies,
- carry out a preliminary assessment of a generic site,
- demonstrate the utility of sensitivity/uncertainty analysis techniques and,
- continue development of the system code.

In addition, a number of technical issues, of either considerable importance to probabilistic analysis in general or required for the Phase II modelling efforts were selected for special study. Among the special topics investigated in Phase II are:

- interpretation of probabilistic results,
- methods for estimating confidence regions on output distributions,
- techniques for handling correlated parameters,
- model response to time-variant input parameters,
- implications of model aggregation.

1.2 Report Structure

Because the Phase II studies covered a variety of aspects of substantially different form and content, and due to the sheer bulk of the material, the report has been separated into three volumes. For example, many of the topical discussion papers are necessarily mathematical in nature and require careful reading with calculator in hand. They have been presented in Volume 2 for readers interested in detailed discussions on specific subjects. In contrast, the system code (excluding the component models) can be described without mathematical equations. The code is discussed in Volume 3 together with the descriptions of the component models which are included as appendices.

In addition, it was also clear that a large amount of new material would be generated through the Phase II investigations. As a consequence it was agreed to structure the Phase II report as follows:

Volume 1:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 1, Main Report.
Volume 2:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 2, Workshop Notes and Topical Discussion Papers.
Volume 3:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 3, Probabilistic Analysis Users Manual and Documentation for the Uranium Tailings Analysis Program (UTAP)

This volume (the main report) has been structured into the following sections:
2.0 Background Considerations
3.0 System Code Structure
4.0 Phase II Probabilistic Assessment Results
5.0 Summary Observations from Topical Discussion Papers
6.0 Summary and Recommendations

REFERENCES

APPENDICES

2.0 BACKGROUND CONSIDERATIONS

The focus of the NUTP modelling work, of which this study is part, is to provide tools which can be used to assess quantitatively the effects of uranium mill tailings. It is realized, however, that natural systems are extremely complex and vary from site to site. Recognition of these factors is reflected in the approach being taken by the NUTP to model development, which is illustrated schematically in Figure 2.1.

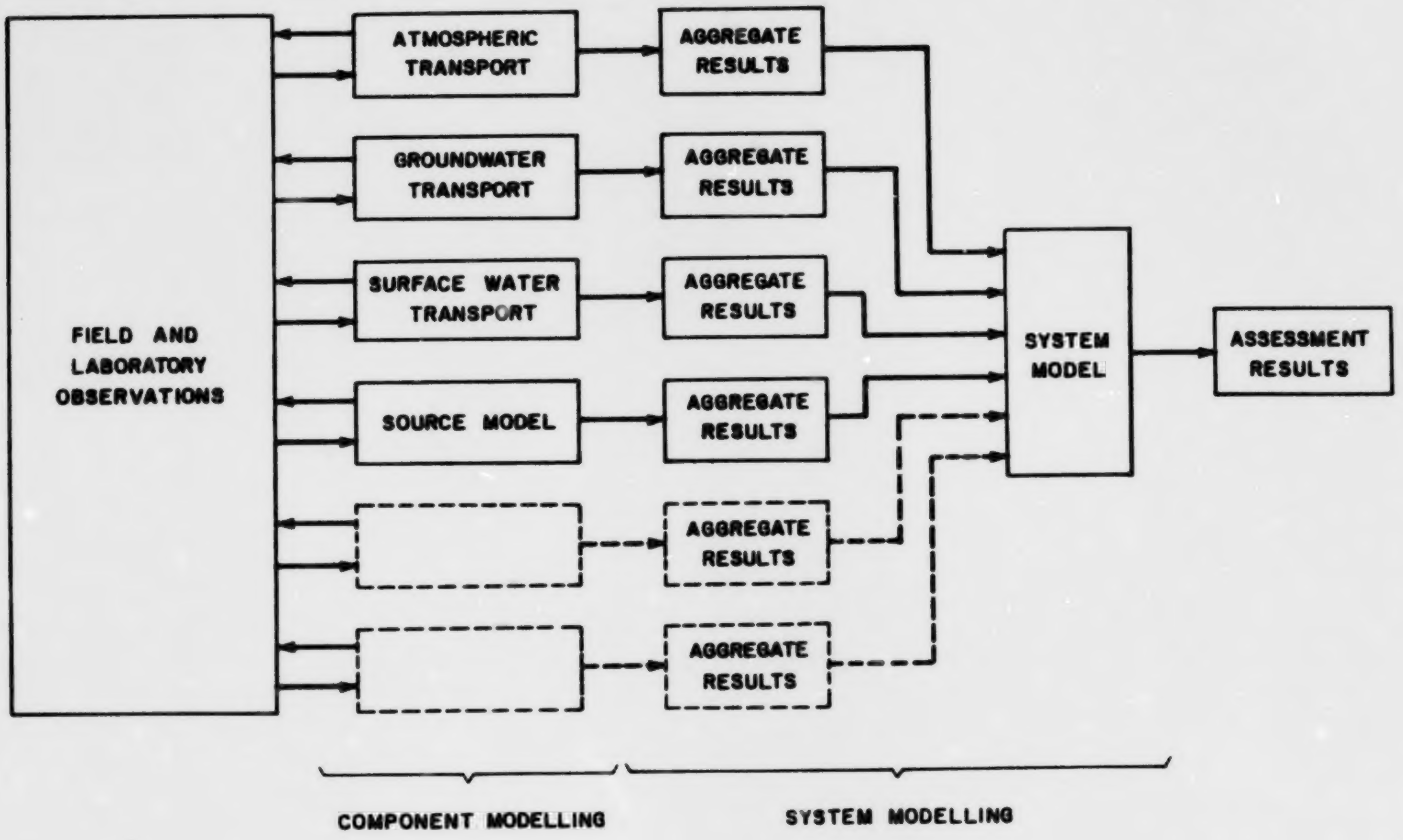
The approach to mathematical modelling involves a series of separate component models which are based as much as possible on field and laboratory observations and contain as much fundamental understanding as is available. The concept is that simplified aggregate models are derived from the detailed component models and that the aggregate models are used in the system simulations. The objective of this approach is to provide a flexible structure which can easily accommodate changes in component models (or aggregate models as appropriate) and model input parameters. At the present level of development the mathematical models used to simulate the tailings system are a mix of quite complex component models (e.g. the tailings source term) and simpler aggregate models (e.g. transfer of radium from water to fish flesh).

Traditionally, the models have been evaluated in a deterministic fashion. Specifically, this means that a single value is selected for each parameter and input to the model(s) resulting in a single estimate of some environmental consequence, such as dose to an individual in the critical group.

Such single value estimates of environmental consequence from deterministic analyses are often challenged since many of the parameter values used as inputs in model evaluations are either incompletely known or inherently variable. Over the past decade,

FIGURE 2·1

NUTP APPROACH TO MODEL DEVELOPMENT



considerable attention has been given to incorporating the uncertainty (and variability) associated with individual parameters directly into the model predictions. This is achieved by a probabilistic approach, in which probability distributions, rather than single values are specified for many of the parameters input to the simulation models. The models are then evaluated repeatedly, with new values randomly selected for the input parameters on each trial. Thus, a probabilistic analysis produces a frequency distribution of the calculated result(s) which can be interpreted in terms of subjective probability. The concept of probabilistic analysis is illustrated in Figure 2.2.

It should be noted that the mathematical models employed in such analyses are usually no different than those formerly used. They are simply being solved in a probabilistic, rather than in a deterministic, manner to take into account known uncertainties. Examples of probabilistic assessment applications are cited in the review presented in the Phase I report (SENES 1984) and include applications to business decision making (Reilly and Johri 1969), evaluation of different weapons systems (Wimmenauer 1971), reactor safety studies (Cairns and Fleming 1977) and analysis of the stability of tailings embankment dams (Utah State University 1982).

In Phase I, the applicability of the probabilistic technique to assessing the effects of uranium mill tailings was demonstrated using a generic reference tailings site, a simplified set of environmental pathways, and assuming quasi-steady conditions. In Phase II, the component models were extended to permit explicit investigation of the time dependence of the system. While the same environmental pathways were used in the mainstream modelling activities in Phase II as were identified in Phase I, additional pathways were investigated in comprehensive

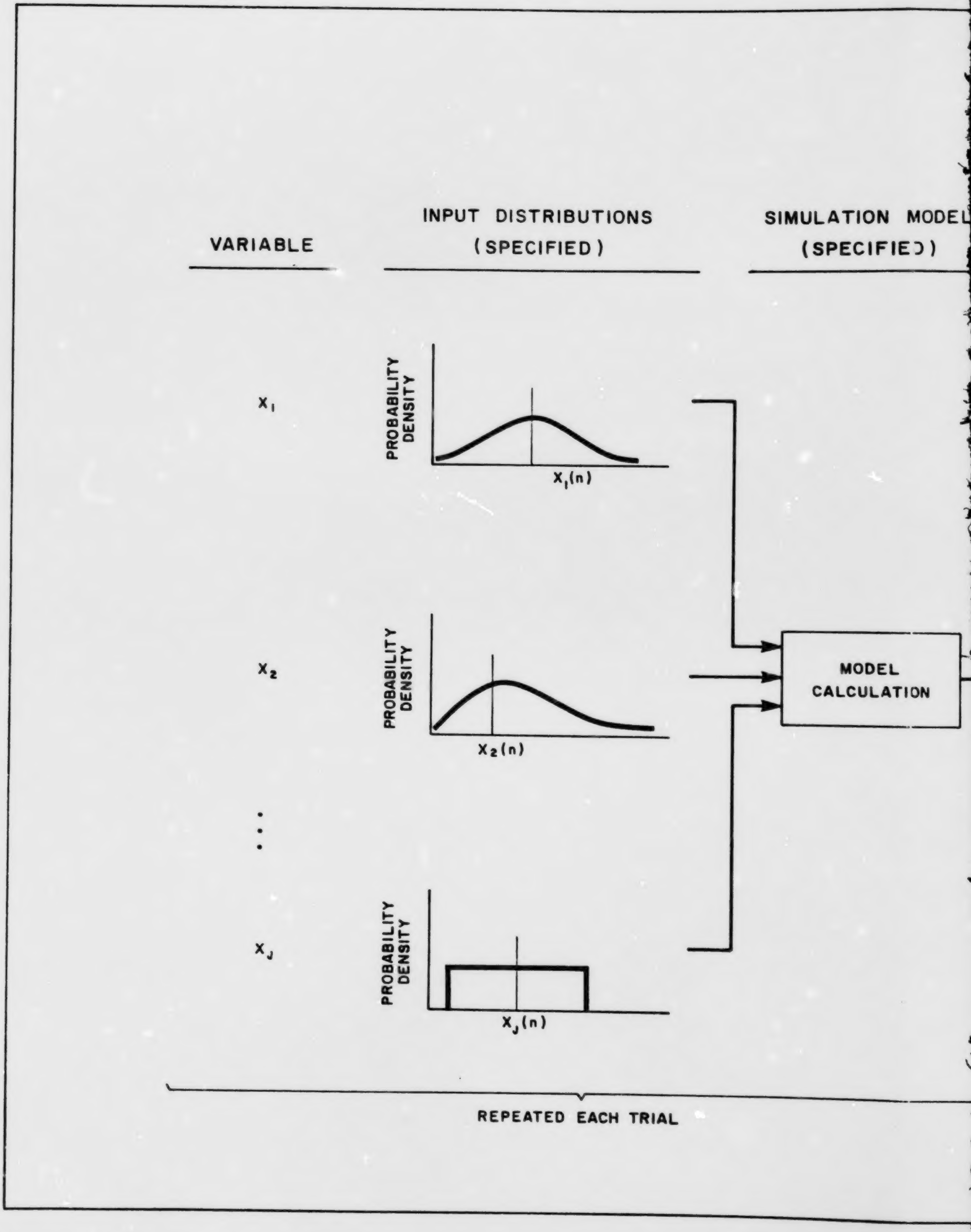
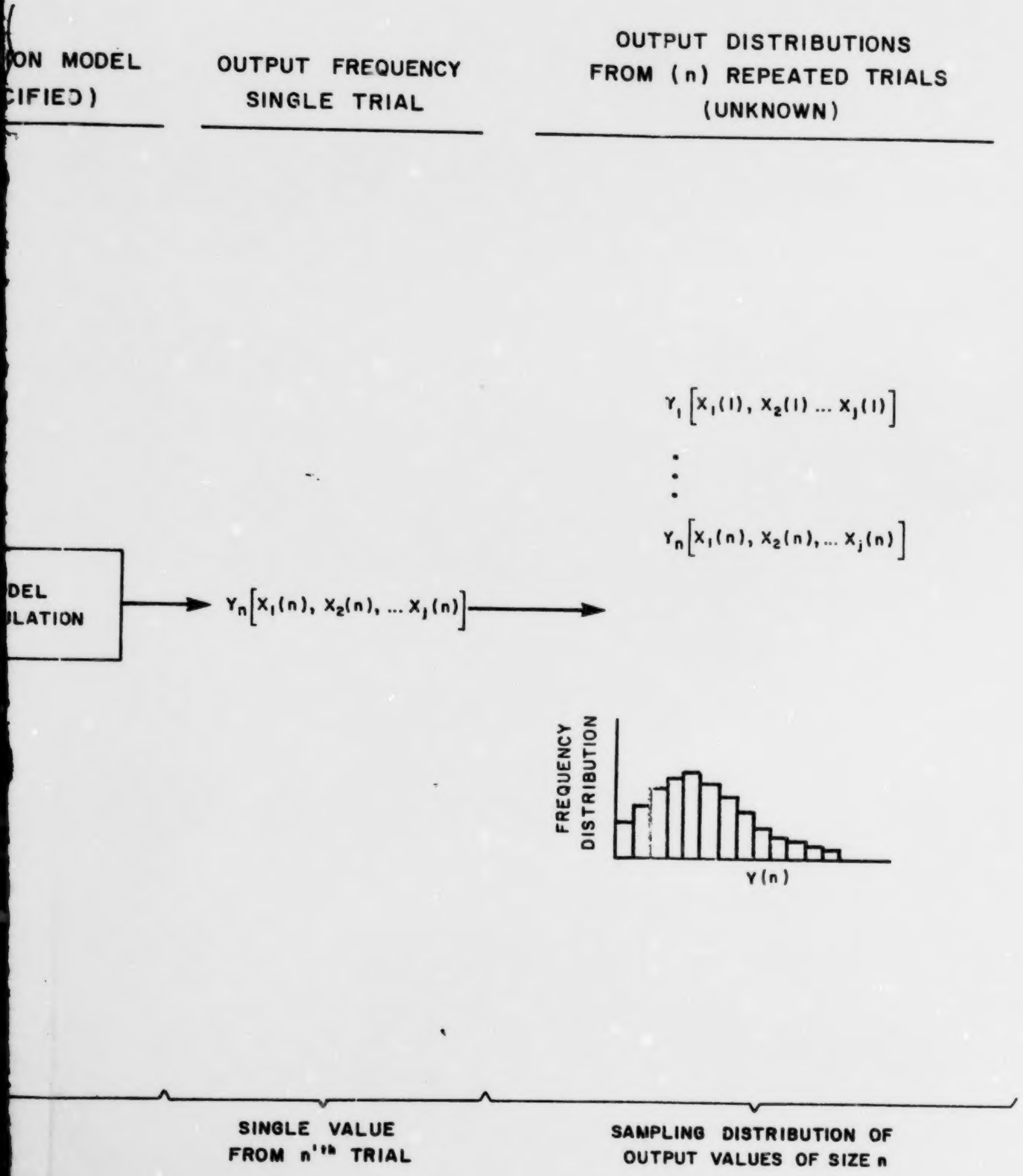

VARIABLE
INPUT DISTRIBUTIONS
(SPECIFIED)
SIMULATION MODEL
(SPECIFIED)
X₁
PROBABILITY DENSITY
X₁(n)
X₂
PROBABILITY DENSITY
X₂(n)
MODEL CALCULATION
X_J
PROBABILITY DENSITY
X_J(n)
REPEATED EACH TRIAL

FIGURE 2·2

CONCEPT OF PROBABILISTIC ANALYSIS



model aggregation studies which were carried out in parallel to the mainstream work. The basic features of the generic reference tailings site described in the Phase I study was retained in Phase II.

This chapter presents background information necessary for an understanding of the Phase II probabilistic assessment, described primarily in Section 4. The reference tailings site, the environmental pathways considered and the reference receptor are all described in this section. In addition to the foregoing, some concepts related to modelling time-dependent behaviour are discussed. Finally, techniques for sensitivity and uncertainty analysis used to examine the results of the Phase II probabilistic assessment are reviewed.

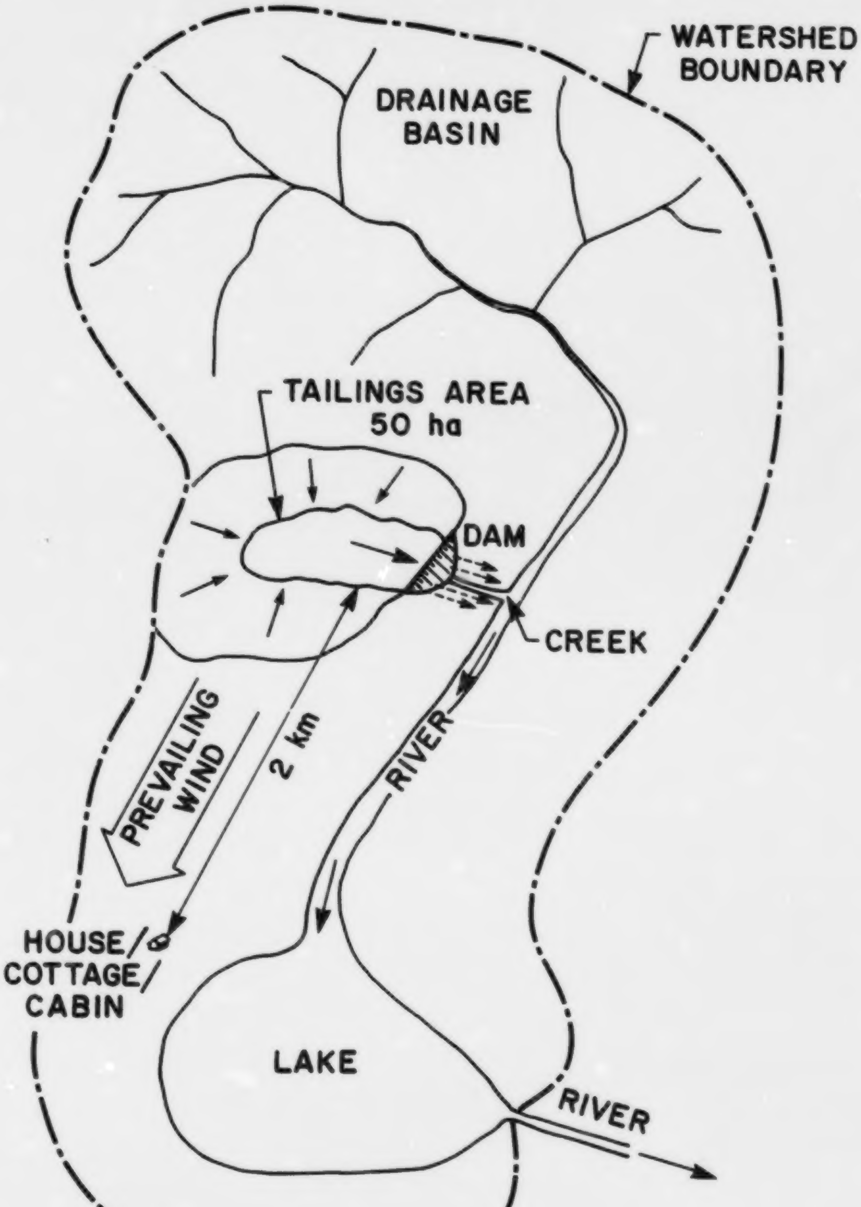
2.1 Reference Tailings Site Characteristics

The basic characteristics of the hypothetical reference tailings site, shown on Figure 2.3, are the same as those assumed in the Phase I studies (SENES 1984). The tailings disposal site occupies a natural depression formed within the upper reach of a valley. The tailings are contained behind a conventional, slightly pervious dam constructed across the low point. Surface and groundwater flows from the site are assumed to follow the valley surface features with ultimate discharge to a major river which flows into a lake.

The reference tailings site is defined to have a surface area of 50 ha and an average tailings depth of 10 m for a total volume of 5×10^6 m^3. The reference tailings have characteristics representative of those resulting from the acid leaching of low-grade pyritic uranium ore containing approximately 0.1% uranium. Ninety-five percent of the uranium is assumed to be recovered during milling, while all of the other radionuclides are assumed

FIGURE 2·3

SCHEMATIC ILLUSTRATION OF SETTING FOR THE REFERENCE TAILINGS SITE



to remain with the tailings. The tailings are neutralized in the mill by lime addition prior to discharge to the tailings areas.

Specifically, the principal characteristics of the tailings are: a particle size distribution of 50% less than 75 um; a pyrite content of 6.5%; a gypsum content of 5%; and, a radium concentration of 5 to 15 Bq.g^{-1} distributed between the tailings solids and gypsum.

Surface runoff from the tailings area (nominally 90% of the net precipitation) is assumed to interact with the surface layer of the tailings and to acquire the chemical and radiological characteristics of the tailings porewater. The surface runoff mixes with river water 200 m downstream of the tailings area. The groundwater flow (the remaining 10% of the net precipitation) moves downward through the tailings pile into a subsurface groundwater aquifer and eventually emerges at the river 200 m from the tailings site.

Both surface water and groundwater flows are diluted with the fresh water flow in the river which eventually discharges to a lake 2 km downstream. The drainage basin for the river-lake system at the outlet of the lake covers 2000 ha. The lake has a surface area of 160 ha and a mean depth of 10 m.

The prevailing wind direction parallels the direction of river flow. Consequently, the reference receptor, located in a cabin on the shore of the lake, is exposed to both air and water pathways.

2.2 The Reference Receptor

The reference receptor was chosen to be an adult male (i.e. Reference Man as defined in ICRP 23 (1975)) who lives year-round in a cabin on the lake, shown on Figure 2.3, 2 km downwind of the reference tailings site. The receptor lives off the land, deriving 50 per cent of his food supply from local sources: eating vegetables and berries grown or harvested near the residence, eating fish taken from the lake, and eating game hunted in the immediate area. All of his drinking water is taken from the lake and is consumed directly without treatment for radium removal.

Assessment of the dose to the reference receptor from radon and radium releases from the reference tailings site was undertaken by pathways analysis. A simplified diagram showing the pathways selected for investigation in Phase II is shown in Figure 2.4. The exposure pathways include:

i) radon source → atmospheric dispersion → inhalation
ii) dust source → atmospheric dispersion → inhalation
iii) dust source → atmospheric dispersion → garden produce → ingestion
iv) dust source → atmospheric dispersion → forage crop → caribou → ingestion
v) water source → surface water transport → ingestion
vi) water source → surface water transport → fish → ingestion

The specific characteristics of, and the factors needed to estimate annual dose to the reference receptor are provided in Appendix F of Volume 3.

A number of additional parameters describing the receptor were assigned probability distributions in Phase II rather than the Phase I constant values. In doing this, however, care was taken to ensure that the variability assigned to the parameters did not change the basic characteristics of the receptor being studied. The parameters treated as variables in Phase II are occupancy factor and fraction of the vegetables and meat produce which are grown locally. These factors would naturally vary from year to year; for example the success in hunting game would have a direct influence on the amount of game consumed in any year. Some factors were not varied, specifically those involving annual intakes of food, water, air and radiation dosimetry.

2.3 Environmental Models

In Phase I the six pathways shown on Figure 2.4 were modelled in order to demonstrate the probabilistic assessment technique and to examine various techniques for sensitivity and uncertainty analysis. Quasi-steady conditions were assumed for this analysis, which included only radium and radon. In addition, the groundwater to surface water pathway was examined and used to illustrate certain features of time-dependent behaviour.

For the Phase II study, it was necessary to modify the component models to permit explicit calculation of time-dependent behaviour of both the component models and the overall system. Also, the component models were expanded to include other selected radionuclides and non-radioactive constituents (in addition to radium and radon).

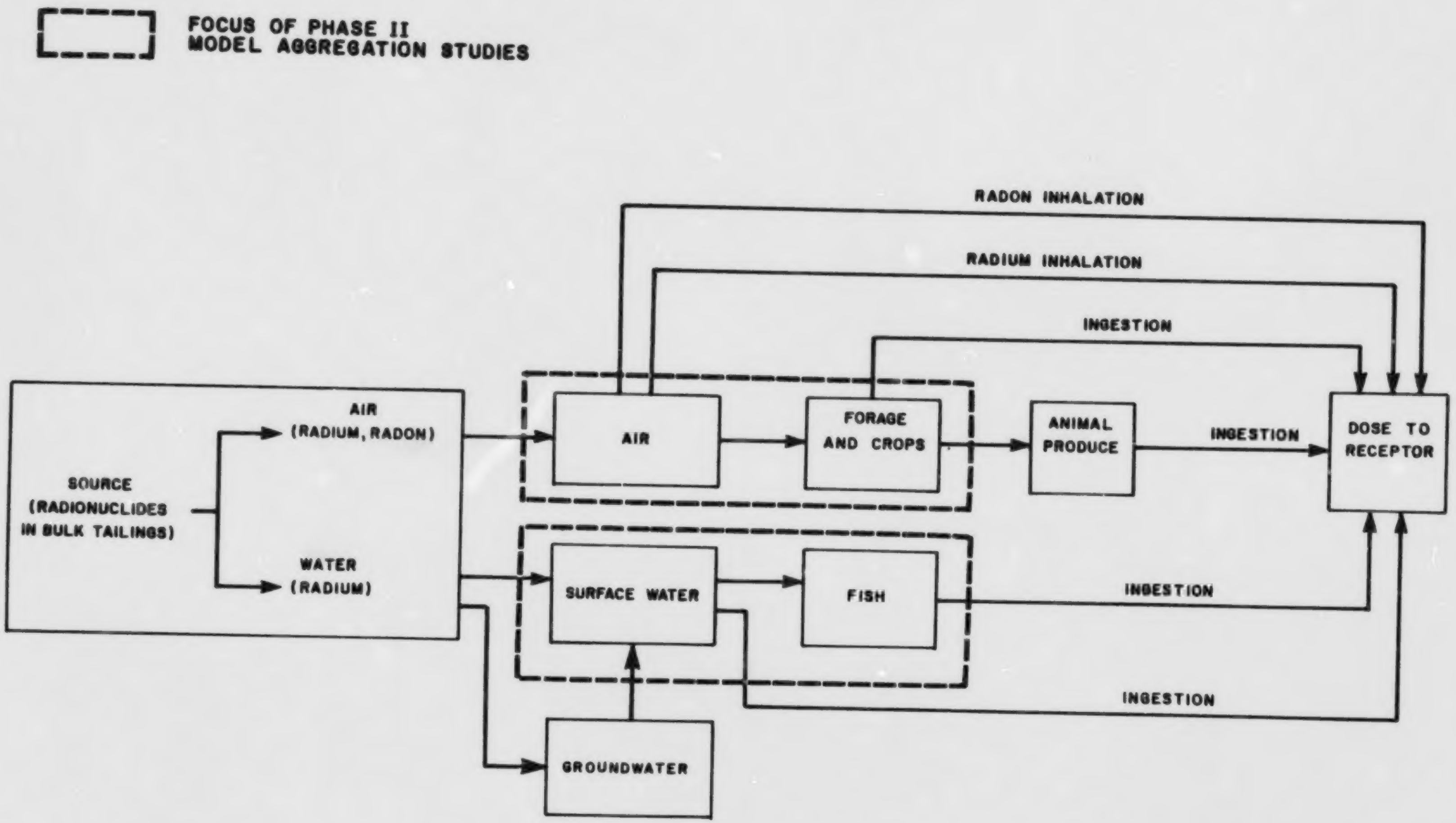
The following section provides a brief description of the essential elements of the Phase II component models. The reader interested in the details of the mathematical formulations as well as a description of the model parameters, is referred to the Appendices of Volume 3.

2.3.1 The Tailings Source Term

Radioactive and non-radioactive constituents from the tailings are released to air and water and are subsequently transferred

FIGURE 2·4

PATHWAYS STUDIED IN PHASE II



to, and accumulate in, the various environmental compartments. Moreover, the rate of release of constituents to the environment changes with time, for example, the rate at which radium leaches from the upper layer(s) of the tailings. This directly affects the air source term, since wind erosion of tailings occurs at the surface and since the radon release rate is directly related to the radium concentration in the uppermost tailings layer(s). Similarly, the time-dependent tailings porewater concentration provides the source concentration for the groundwater and surface water models. Consequently, the biochemical and physical processes, which describe the behaviour of radium within the tailings mass, determine to a large extent the overall time-dependent behaviour of the system. The water and air components of the tailings source term are discussed below.

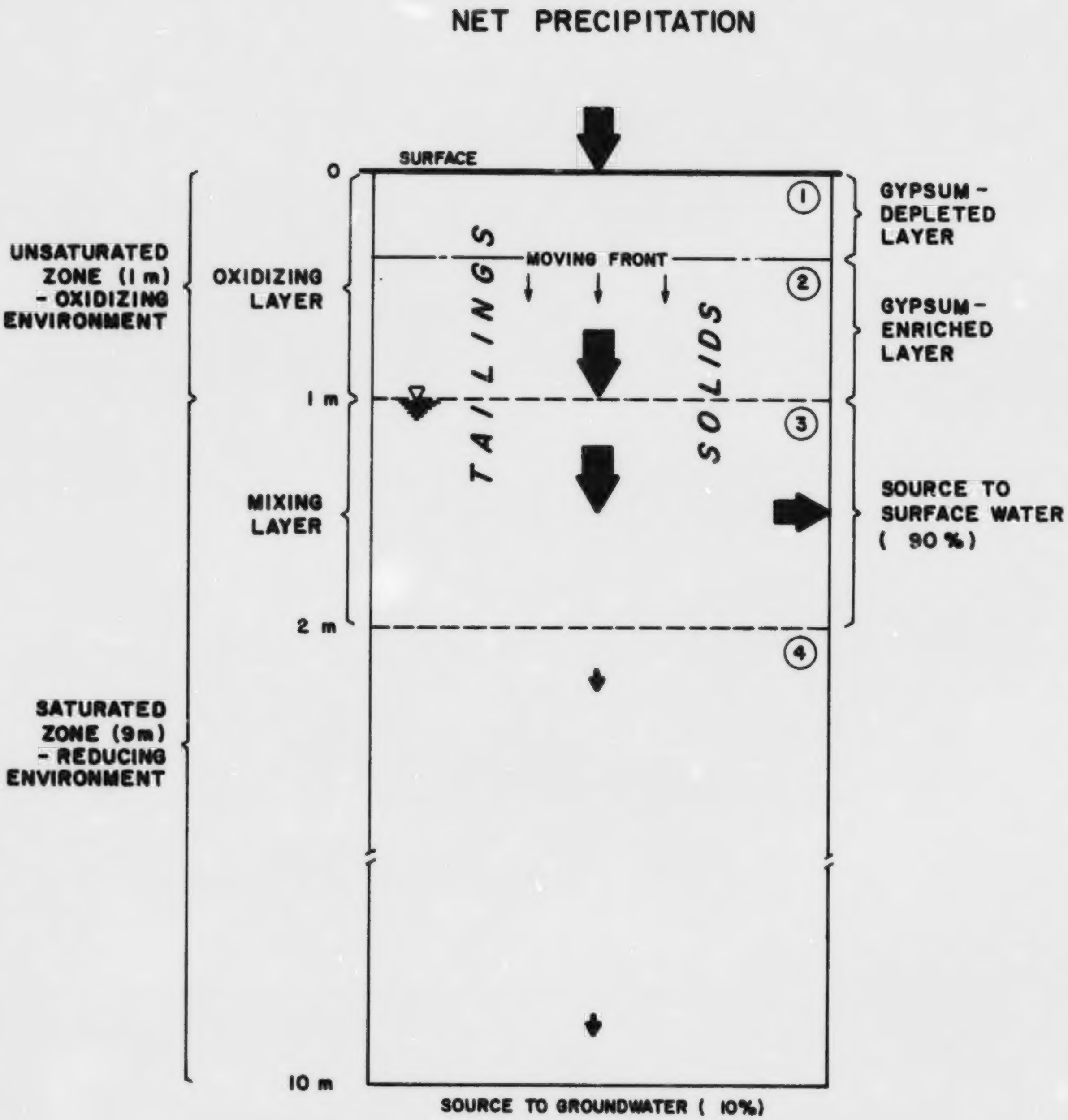
The Water Source Term

In Phase II, the tailings source is modelled as a layered system, as shown in Figure 2.5. The features shown on the figure are explained in detail in Appendix A of Volume 3. The species modelled within the water source term (i.e. in the tailings porewater) include calcium, sulphate, iron, radium and thorium. The pH of the tailings porewater is also calculated based on ionic equilibrium considerations.

As indicated on Figure 2.5, the tailings are assumed to consist of an upper, unsaturated (oxidizing) zone, 1 m deep. The remaining 9 m below the interface is assumed to be saturated. The unsaturated zone is further subdivided into two layers, a gypsum-depleted layer and a gypsum-enriched layer. In the uppermost layer of the unsaturated zone, gypsum dissolves in the excess precipitation which penetrates the tailings surface and moves downward to the water table. This effect is depicted on the figure as a moving front which separates the gypsum-depleted

FIGURE 2·5

CONCEPT ADOPTED
FOR MODELLING THE TAILINGS SOURCE TERM


NET PRECIPITATION
SURFACE
0
1 m
2 m
10 m
TAILINGS
SOLIDS
MOVING FRONT
①
②
③
④
UNSATURATED ZONE (1 m) - OXIDIZING ENVIRONMENT
OXIDIZING LAYER
MIXING LAYER
SATURATED ZONE (9m) - REDUCING ENVIRONMENT
GYPSUM - DEPLETED LAYER
GYPSUM - ENRICHED LAYER
SOURCE TO SURFACE WATER (90 %)
SOURCE TO GROUNDWATER (10%)

and gypsum-enriched layers. The concentration of dissolved gypsum in the porewater depends on the concentration of sulphate in solution. As the sulphate level increases, the dissolved gypsum level decreases. Since the sulphate level increases with depth in a tailings pile as a result of pyrite oxidation, gypsum reprecipitation occurs as the porewater moves downward. This concept is depicted by the gypsum-enriched layer on Figure 2.5.

Since radium exists as a co-precipitate with gypsum, it follows that radium will dissolve and reprecipitate in concert with gypsum. Hence, the dissolved radium levels in the porewater will parallel that of gypsum. In addition, the source term model applied in Phase II accounts for radium-226 ingrowth from the decay of thorium-230. The behaviour of thorium and thorium-derived radium is also considered.

The upper 1 m of the saturated zone has been specified for modelling purposes as a 'mixing layer' in which adsorption of several species may occur. This feature is considered necessary to account for the natural retardation processes which occur in large tailings areas. The radium concentration in the porewater in the mixing zone is calculated as a direct function of the liquid-to-solid phases distribution coefficient, K_D. Adsorption on organic and inorganic solids and chemical precipitates has been found to retard significantly the movement of radionuclides. The thorium concentration in the mixing zone is calculated as a function of pH as described in Appendix A of Volume 3. In addition to the above, the model takes into account ferric iron reduction to ferrous iron in the mixing zone, and the attendant pH shift which occurs.

Porewater pH is modelled as a function of time using an equilibrium model based on a quartz/kaolinite/ferric hydroxide/gypsum/sulphuric acid system. In the mixing zone, pH is calculated explicitly as a function of iron, calcium and sulphate.

Seepage entering the surface water flow regime (i.e. the surface water source term), is assumed to have the characteristics of porewater in the (homogeneous) mixing layer. Similarly, the mixing layer is the source of recharge to the deep saturated tailings zone which feeds the groundwater flow regime beneath the tailings pile.

Air Source Terms

In Phase II an empirical model, based on experimental work at mining operations in the United States, was used to simulate the dust source term (CDOH 1984). The key input parameters to the dust source model are: the silt content of tailings (actually the fraction of tailings of erodible size), the percentage of time that the wind speed exceeds a threshold velocity for wind erosion ($5.4\ m.s^{-1}$) and the number of days on which there is measurable precipitation (>0.25 mm). A conservative approach was used to estimate the meteorological conditions for the dust source model. The two meteorological parameters (wind speed and precipitation) were derived from observations of a representative northern Ontario monitoring station, taken during the six summer months. These conditions were assumed to prevail for the entire year. In actuality the dust suspension during the winter months is likely to be minimal, hence the model may overestimate this quantity by as much as a factor of two. Ongoing NUTP research in this area may be able to guide improvements in model formulation and/or parameter selection.

The airborne radionuclide release rate decreases with time as the radioactive constituents are leached from the tailings surface layer (see previous discussion).

The approach used in Phase II to model the release of radon-222 considers the tailings to consist of an unsaturated layer (the source of radon), and a saturated layer. The saturated layer does not act as a source of radon because the water effectively inhibits its diffusion.

The unsaturated layer is divided into the two zones described by the water source model: an upper gypsum-depleted zone (i.e. a zone of reduced radium content) and a lower gypsum-enriched (i.e. radium-enriched) zone. The depth of the interface between the two zones increases with time until the entire unsaturated layer is gypsum-depleted. The behaviour of the interface is determined by the water source term and depends, in a complex manner, on the chemistry of the system.

In terms of radon release to the atmosphere, the two zones are considered to be homogeneous, with each zone characterized by an average radium concentration, moisture content, porosity, etc. This model for radon release is equivalent to a covered tailings model in which the tailings are of finite thickness and the cover also contains radium, i.e. the "cover" is the upper zone and the "tailings" are the lower zone. The theoretical development of such a model was carried out by Rogers et al (1980).

More details on the air source models can be found in Appendix B of Volume 3.

2.3.2 Environmental Dispersion

Dispersion in the Surface Water

Contaminated surface and groundwater flows from the tailings area are intercepted in a river which discharges into an oligotrophic lake with a surface area of 160 ha and mean depth of 10 m. (An oligotrophic lake is a low productivity lake as measured by biomass or some other indicator of biological activity.) The drainage basin at the outflow of the lake equals 2000 ha as compared to the tailings area which covers 50 ha. The waters of the drainage basin have a naturally low buffering capacity, low nutrient content and low total dissolved solids level.

The estimation of pH and constituent concentrations in the reference lake takes into account physical, chemical and biological factors. For example, radium rapidly adsorbs onto the surface of organic and inorganic matter and is removed from the aqueous phase with sedimenting solids. Thus, radium behaviour in the lake is modelled on a mass basis to permit removal from the water column and feedback from the sediment to be followed through time. A detailed description of the concepts of the lake model is provided in the discussion paper by Scharer and Halbert (Appendix F of Volume 2).

The major water quality constituents, including calcium and sulphate, are modelled as conservative substances. The concentration of these constituents are affected primarily by dilution; taking into account natural background levels in the fresh water system. Lake pH is calculated assuming ionic equilibrium. Species taken into account in the lake pH calculations include iron, aluminum, calcium, sulphate, sodium, potassium and magnesium levels, as well as dissolved carbonate (in equilibrium with CO_2 in the air).

Further details on dispersion in the water environment are provided in Appendix C of Volume 3.

Air Dispersion

The gaussian plume atmospheric dispersion model used in the Phase I study (SENES 1984) was preserved in its entirety. The basic form of the model is outlined in Appendix D of Volume 3.

2.3.3 Environmental Transfer Models

The models used to calculate the concentrations of radionuclides in various environmental media are essentially those used in the Phase I studies (SENES 1984). The following models are described in Appendix E of Volume 3.

- concentration of radionuclides in the edible portion of vegetation due to direct deposition from the atmosphere;
- concentration of radionuclides in animal produce due to vegetation ingestion; and,
- concentration of radionuclides in edible portions of fish.

Unlike Phase I, however, it was considered desirable to assign distributions to some parameters which had previously been given as nominal (and in some cases conservative) values (SENES 1984). This not only adds an element of realism to the modelling exercise but also permits these parameters to be included in sensitivity and uncertainty analyses, thus permitting assessment of the importance of the parameters.

2.4 Model Aggregation Studies

There is considerable demand in assessment modelling to account explicitly for distinct physical and biological processes, even when the data are available to permit the use of aggregated or "lumped" parameters that supposedly account for these processes implicitly. Therefore, model aggregation studies were carried out to investigate the defensibility of replacing complex component models by simpler aggregate models.

Two specific environmental components indicated by the dashed areas on Figure 2.4 were selected to examine the effects of model aggregation, namely: a) the prediction of the steady-state concentration of Ra-226 in the edible tissues of fish as a consequence of a prolonged release of this radionuclide into a lake; and b) the prediction of the steady-state concentration of Ra-226 in pasture forage (grasses) as a consequence of a continual deposition of radionuclides from the atmosphere to vegetation and soil.

These subjects are examined in detail in the discussion papers by Hoffman, Gardner and Blaylock found in Volume 2. Summary observations from these investigations are also provided in Section 5 of this volume.

2.5 Disposal Technologies Considered

The reference tailings area for the Phase II study contains the tailings from a pyritic, low-grade uranium ore which has undergone sulphuric acid leaching in the mill. The tailings are comparatively coarse (approximately 50% retained on a 200 mesh screen) and are highly acid-producing.

Although there have been numerous technologies suggested for long-term management of uranium mine tailings, few have been shown to be economically viable for this type of low grade deposit located on the Canadian Shield. Perhaps the most promising technology is the application of a surface cover to reduce and control the rate of acid production in the surface oxidizing zone. The benefit of reducing the acidity of the tailings porewater is to minimize the rate of solubilization of numerous heavy metals and radionuclides.

The preferred cover would be a local soil. However, due to the scarcity and high cost of suitable borrow materials, this cover option is neither practical nor economical in most instances. The next most logical surface cover material would involve application of a suitably treated layer of tailings solids. Several potential alternatives exist including:

- limestone incorporation into the surface zone
- depyritization of the surface zone (via flotation)
- depyritization and radionuclide reduction (via bulk flotation).

Limestone addition involves the placement of a tailings cover with limestone incorporated in stoichiometric proportion to the pyrite content. The limestone will act to inhibit bacterial activity and neutralize any acidity produced. Limestone addition has proven to be successful in providing pH conditions suitable for the vegetation of the acidic, non-operating tailings areas. Therefore, vegetation of the surface would logically be implemented to control wind and water erosion. For operating tailings areas, methods of limestone addition have not been developed, however it is believed this could be achieved by adding crushed limestone directly into the tailings discharge. The primary concerns with this technology are the depletion of limestone before all the pyrite is oxidized, or the reduction in limestone solubility due to blinding of the surface by metal hydroxides.

The depyritization concept would involve the placement of a cover of pyrite-reduced tailings on the surface of a tailings area. Experimentation has shown that 90 to 95 percent of the pyrite can be removed by flotation. Depending on the nature of the flotation reagents, radium may or may not also be concentrated.

The pyrite-reduced tailings contain little or no limestone and are therefore subject to acid production. Oxidation experiments on depyritized tailings have indicated that the residual pyrite would be oxidized from the surface zone in a comparatively short period of time. Without limestone incorporation, active surface water treatment to control acid runoff would be required during this period. With limestone incorporation, acidity would be controlled and pyrite would oxidize at a somewhat slower rate since only alkaline chemical oxidation would be anticipated to occur.

Bulk flotation is a process developed by CANMET which involves removal of much of the pyrite during the milling process (following uranium recovery). The floated material (or concentrate) comprises approximately 25 percent of the material originally destined for tailings. This includes 90 to 95 percent of the pyrite and 60 to 70 percent of the radium and thorium. As with depyritization, limestone incorporation would be included, if required, to neutralize acid produced in the short term.

The most promising of the surface cover options would appear to be either pyrite flotation or the combined pyrite plus radionuclide bulk flotation process. The options selected for the modelling investigations thus focussed on the base case scenario (no special action) described in Phase I and three alternative disposal scenarios which employ a combination of flotation technologies and limestone addition.

The chemical and radiological characteristics of the surface cover zone (i.e. the depth of the surface cover material) for the base case and three disposal alternatives are summarized in Table 2.1. It has been assumed that the cover material would be applied to a nominal depth of 1 m. This would have the benefit

Table 2.1

Chemical and Radiological Characteristics of Uranium Tailings Surface Zone (Upper 1 m)

Scenario	FeS_2 %	$CaSO_4$ %	$CaCO_3$ %	Ra-226 Bq/g	Th-230 Bq/g
1 - Base Case	6.5	5	0	11.1	11.1
2 - Pyrite Reduction	0.6	5	0	11.1	11.1
3 - Pyrite Reduction and Limestone	0.6	5	1.0	11.1	11.1
4 - Bulk Float and Limestone	0.6	0.5	1.0	3.3	3.3

of minimizing oxygen penetration into the pyritic tailings zone essentially eliminating acid production and metal leaching.

The major difference between the base case and the alternative schemes is that the 1 m deep surface oxidizing zone has been replaced with pyrite-reduced and limestone-treated tailings cover material. Further details regarding the alternative scenarios are presented in Appendix A of Volume 3.

2.6 Some Aspects of Time Dependency

The discussion paper by Beals et al (Appendix D) in Volume 2 discusses various ways of treating systems which evolve in time and parameters which vary with time. Several aspects of time dependency are also discussed in the paper by Kaplan et al (Appendix A) in Volume 2.

First, consider the deterministic evaluation of a time-dependent system. Suitable parameter values are entered into the mathematical model of the system and values of the time-dependent output variables(s) are calculated as the system evolves in time from a specified initial state. This approach may also be taken in a probabilistic evaluation of the model. In this case, however, the actual parameter values are not uniquely specified, but rather for each trial (i.e. each individual simulation) a set of parameter values are selected by randomly sampling from a suitable distribution of possible values for each parameter. The system then evolves through time with the parameter values fixed at the initially-selected values. In the next trial, new parameter values are selected by randomly sampling again from the specified parameter distributions, and the system is allowed to evolve once more in time. The process is repeated until the desired number of trials have been carried out.

In essence this procedure produces a family of (time) curves, one curve for each trial. Each curve represents the behaviour of the time-dependent variable with respect to the particular set of randomly-selected input parameters. Consider now a specific variable; for instance, total dose. Each trial will produce a different curve of total dose versus time. At any specific time the actual values calculated for the total dose may differ widely from one trial to the next. Values at a particular time from all of the trials can be used to construct a frequency diagram and calculate sample statistics such as the mean and variance. The "sample" in this instance is the ensemble of individual values of total dose obtained from the same time interval for each trial. Examples of this are presented in Section 4.2.2 (Figure 4.7).

Now consider a particular time-variant input parameter such as the quantity "mean annual wind speed". The value of this parameter in any past year is known (in the sense that it can be estimated from observational data) and is a single constant value (except for observational uncertainty). However, the mean annual wind speed varies from year to year and hence, even if there were no observational uncertainty, the wind speed in any future year would be inherently uncertain, in the sense that it would vary from year to year. Variability from year to year in input parameters also gives rise to uncertainty in the output variable.

In summary, uncertainty in model prediction(s) may arise from a lack of knowledge of the constant input parameters or from a lack of ability to predict the time-variant parameters. Uncertainty of both sorts is propagated through the mathematical model and is inherent in the output distribution(s). In either case the output distribution(s), which can be interpreted in terms of subjective probability, describes how strongly one is entitled to believe in a particular consequence in a particular year.

In the Phase II models some of the input parameters are not only uncertain but vary from year to year. Several alternative approaches to incorporating such uncertain, time-varying parameters were outlined and discussed in the paper by Beals et al. in Volume 2. The techniques are briefly discussed below:

1) where a time increment of one year is used in the simulation each time-variant parameter is sampled directly from the distribution of annual values at each time increment;

2(a) where a larger time increment is used in the simulation, the distribution of mean annual parameter values requires adjustment. The approach taken in this study was based on consideration of the sampling distribution of the mean. Consider, for example, mean annual windspeed to be a random variable (w) with a population mean of μ_w and a variance of σ_w^2 and that $\overline{w}$ is the sample mean from a random selection of n values of w and S_w^2 is the variance of the same sample. Because they vary from one sample to the next, the sample statistics are themselves random variables with the following properties:

$$\begin{aligned} E(\overline{w}) &= \mu_w \\ \mathrm{Var}(\overline{w}) &= \sigma_w^2/n \\ E(S_w^2) &= \sigma_w^2 \end{aligned}$$

Now, step through time in n-year increments. The sampling distribution of mean annual windspeed would, for practical purposes, be normal with a mean μ_w and a variance of σ_w^2/n. In terms of our original sample of values of w, the sampling distribution of the mean is specified by the mean $\overline{w}$ and the standard error of the mean: $SE(\overline{w}) = S_w/\sqrt{n}$.

2(b) time-varying parameter distributions can be adjusted to the entire simulation period by using the same method discussed in 2(a). In this case the standard deviation of the annual distribution is divided by the square root of the simulation period. For each trial a single value is sampled from this distribution. The concepts in 2(a) and (b) are illustrated on Figure 2.6.

The procedures outlined in 2(a) and (b) are equivalent in concept to constructing new distributions with n-yearly mean values from available data of one year means.

3) time-varying parameters (e.g. precipitation) can also be treated as constants, set at their (long-term) mean values.

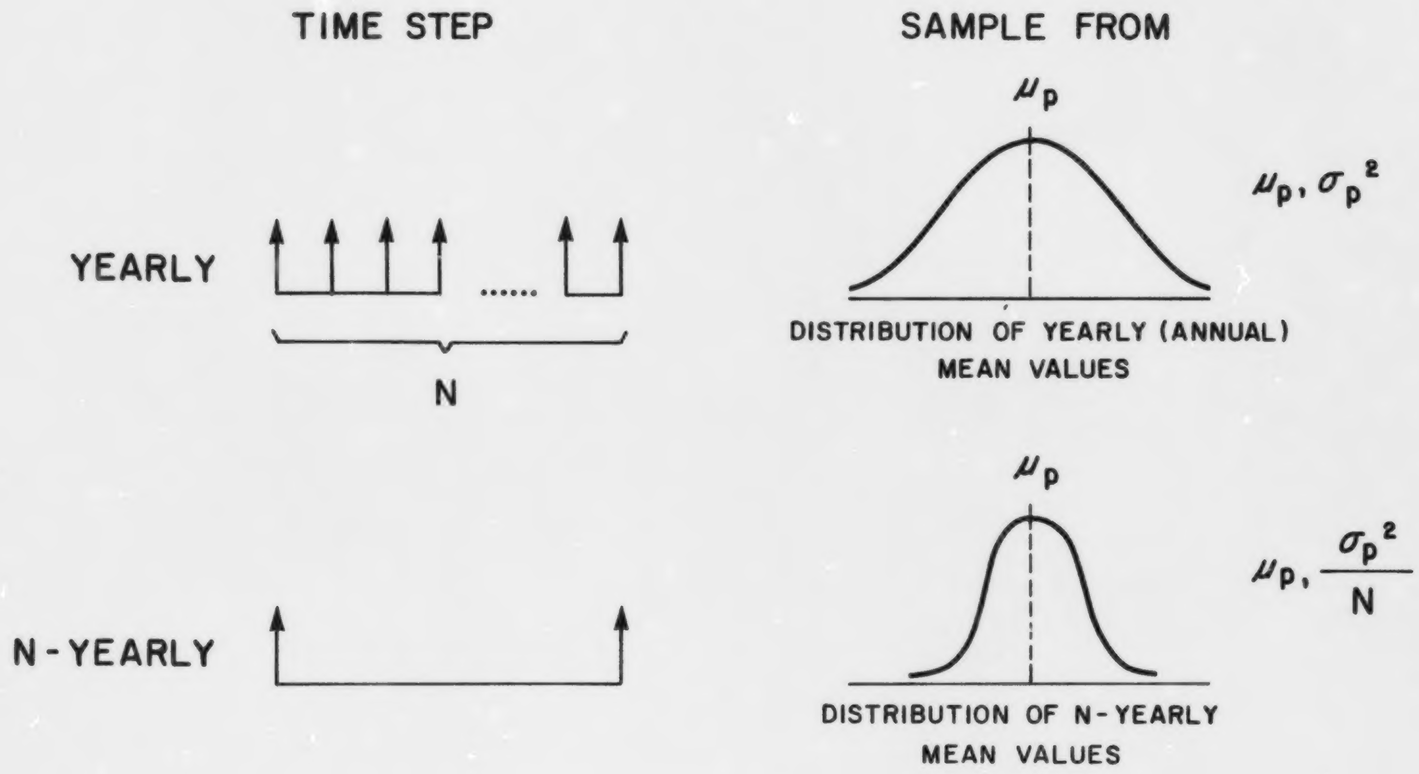
The technique outlined in 2(b) above was used for the modelling in Phase II, in part to facilitate the inclusion of time-varying parameters in sensitivity/uncertainty analysis. A fuller discussion of the implications of time-varying parameters on the identification of important parameters is contained in the section on sensitivity/uncertainty analysis in Section 5. In addition, the reader is referred to the discussion paper by Beals et al. in Volume 2.

2.7 Sensitivity and Uncertainty Analysis Techniques

In probabilistic modelling, the terms sensitivity and uncertainty analysis refer to techniques which seek to define the relationship(s) between model inputs and model predictions. The term "sensitivity analysis" refers to establishing the sensitivity of the magnitude of the model output to changes in the magnitudes of the input parameters. The term "uncertainty analysis" refers to establishing how much the uncertainty (or spread) associated with each parameter contributes to the total uncertainty (or spread) of the model predictions. Parameters to

FIGURE 2·6

CONCEPTUAL APPROACH
TO SAMPLING TIME - VARYING PARAMETERS


TIME STEP
SAMPLE FROM
YEARLY
N
μ_p
μ_p, σ_p^2
DISTRIBUTION OF YEARLY (ANNUAL) MEAN VALUES
N - YEARLY
μ_p
$\mu_p, \frac{\sigma_p^2}{N}$
DISTRIBUTION OF N - YEARLY MEAN VALUES

which model predictions are sensitive and/or parameters which contribute an appreciable fraction of the uncertainty in the model predictions are called "important" parameters in this report. The use of sensitivity and uncertainty analysis techniques to isolate and/or rank the most "important" parameters, i.e. those which have the greatest influence on the model predictions, is referred to as "importance" analysis.

Several techniques for sensitivity and uncertainty analysis were identified in the Phase I studies; the simplest of which were applied to the Phase I probabilistic assessment (SENES 1984). The trend in sensitivity and uncertainty analysis towards the use of the more sophisticated tools of statistical regression was also noted, but application of regression techniques to probabilistic analysis results was beyond the scope of the Phase I study. In Phase II, the investigation of sensitivity and uncertainty analysis was extended to include multiple regression analysis, on either raw or transformed data. Our investigations of the various techniques for sensitivity and uncertainty analysis indicate that no single method is ever likely to be best always.

Two sensitivity and uncertainty analysis techniques were selected for use in the main Phase II studies, namely: stepwise regression and the sensitivity index. Stepwise regression is a relatively complex method of ranking parameters which has considerable statistical support, while the sensitivity index is a simple, more intuitive method. The combined application of these two techniques to the output of the Phase II simulations is referred to as importance analysis and is discussed in Section 4. A brief discussion of the two techniques follows.

A detailed description of several sensitivity and uncertainty analysis techniques is contained in the discussion paper by Beals et al (in Appendix D of Volume 2.) The reader interested in further discussions of sensitivity and uncertainty analysis techniques beyond those provided in this report is referred to recent articles such as those by Iman and Helton (1985), Downing et al (1985), Gardner (1985), Hoffman and Gardner (1983), Hendrickson (1984), INTERA (1983), and Rose and Swartzman (1981). For a fuller discussion of regression analysis the reader should refer to Draper and Smith (1981) and Edwards (1979).

2.7.1 Sensitivity Index

The sensitivity index (SI) (Hoffman and Gardner 1983) provides a simple indication of the response of the model to changes in the input parameters. It is determined by evaluating the model deterministically twice for each input parameter, once with the parameter of interest at its extreme high value and once with the parameter at its extreme low value. All other parameters are held constant at their nominal values. The sensitivity index (for the ith parameter) can then be calculated as:

$$SI_i = 1 - \frac{\text{low calculated result}}{\text{high calculated result}}$$

The closer the SI is to one (i.e. the larger the difference between the high and low calculated result) the more sensitive the model is to the variable. The direction of the model response, that is, whether the response varies in direct or inverse proportion to changes in parameter values (i.e. whether the low or high input parameter value produces the low (or high) calculated result) is also of interest. This is termed the "influence" and is designated "+" or "-" on the parameter importance tables presented in Section 4.

The sensitivity index must be used with caution since the results will vary depending on the nominal values chosen for the other parameters while examining the effects of one particular parameter. It is also implicitly assumed that the model responds linearly in a monotically increasing or decreasing fashion over the full range of input parameter values being evaluated. Notwithstanding these limitations, the SI was found to be applicable to all situations and scenarios in which it was used in Phase II.

2.7.2 Stepwise Regression

The sensitivity index just described applies (rigorously) only for one set of input parameter values (i.e. at only one point on the "model response surface"). Regression analysis, on the other hand, provides an indication of parameter importance while all parameters are allowed to vary simultaneously over their specified ranges. This type of analysis treats probabilistic results as one would treat the results of many trials of a given experiment. The input parameters represent the independent variables while a calculated output (eg. total dose to reference receptor) is selected as the dependent variable.

Linear regression involves the construction of a linear, predictive equation (using all the independent variables) which accounts for as much of the variation in the dependent variable as possible. The overall correlation coefficient (R^2) provides an indication of how much of the variation in the dependent parameter has been accounted for by the given regression equation.

Stepwise linear regression differs somewhat, in that the independent variables are added to the regression equation one-at-a-time, in order of their statistical importance. The order of entry is based on the relative magnitudes of the "partial correlation coefficients" at each step in the analysis. Thus, a new regression equation, and its associated R^2 value are produced at each step in the analysis.

It is easy, in the case of regression analysis to get mired in technical details. However, one must not lose sight of the original objective of the exercise, to identify the important parameters. To emphasize this point, a simplified approach has been taken in presenting the results of stepwise regression.

In this study the parameters were ranked in order of their entry into the regression equation, the most important parameter entered first and so on. Statistically, this order of entry, and hence of parameter ranking is based on the relative magnitudes of the partial correlation coefficients at each step in the regression. For presentation purposes the increase in the overall regression coefficient (R^2) as each parameter was entered in the regression equation was considered to provide a more intuitive indication of parameter importance and hence was provided in the Section 4 on parameter importance (see Appendix A of this volume) instead of the partial correlation coefficients. Further explanation and details of the technique are provided in the paper by Beals et al in Appendix D of Volume 2.

3.0 SYSTEM CODE STRUCTURE

3.1 Overview of the Systems Approach

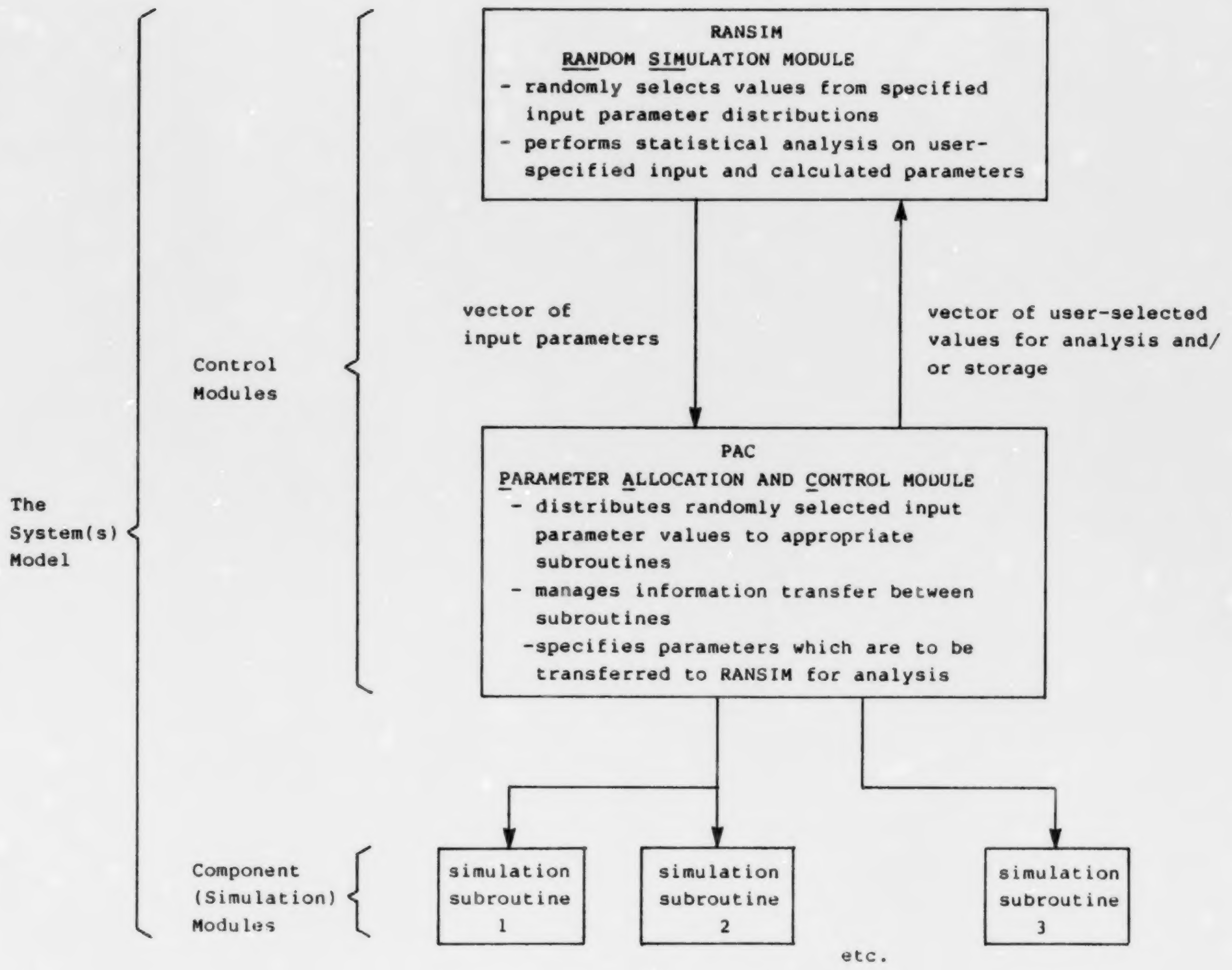
The probabilistic analysis process is inherently large and complex, hence a logical, step-by-step approach is required to make it tractable. The process is exceptionally well suited to the application of a "top-down" computer programming approach. The top-down approach first involves identification of the overall system requirements, followed by the specification of large, conceptual structures consistent with these requirements, followed in turn by further development of conceptual substructures. The overlying conceptual structures always dictate the nature of lower level structures. The requirements and specifications are subsequently translated into identifiable tasks. Once the conceptual algorithms are developed in this logical, top-down manner, implementation is often a fairly mechanical programming exercise.

By way of example, let us consider the overall requirements for performing a probabilistic analysis. We require the ability to draw random samples from specified probability distributions, the ability to route these values to appropriate, user-specified simulation models, and the ability to store and/or analyse the large quantity of results generated by a probabilistic analysis. The input values must be sampled from the appropriate distributions, and, if several simulation models are being used, they must be evaluated in the correct order and any intermediate, calculated values must be passed between them correctly.

Given these overall requirements, the systems structure shown on Figure 3.1 was developed. As indicated, the system model is composed of "control modules" and "simulation (component) modules". The latter are the mathematical models describing the

FIGURE 3·1

CONCEPTUAL FLOW DIAGRAM
OF SYSTEMS APPROACH


RANSIM
RANDOM SIMULATION MODULE
- randomly selects values from specified input parameter distributions
- performs statistical analysis on user-specified input and calculated parameters
vector of input parameters
vector of user-selected values for analysis and/or storage
PAC
PARAMETER ALLOCATION AND CONTROL MODULE
- distributes randomly selected input parameter values to appropriate subroutines
- manages information transfer between subroutines
-specifies parameters which are to be transferred to RANSIM for analysis
simulation subroutine 1
simulation subroutine 2
simulation subroutine 3
etc.
Control Modules
Component (Simulation) Modules
The System(s) Model

system of interest. They are, in general, unique to each problem.

The primary control module (RANSIM), on the other hand, is a completely general computer program designed to perform probabilistic analysis of any suitably specified problem. As an interface between the completely general probabilistic "tool" and the many simulation modules required in this analysis, an intermediate control module was developed. The Parameter Allocation and Control (PAC) module, indicated on the figure, fulfills this intermediary role. The conceptual attributes of PAC are generally the same for different component modules, however, the details of implementation will change.

In keeping with the top down approach, each conceptual module in the overall systems approach will be described in more detail in the following sections.

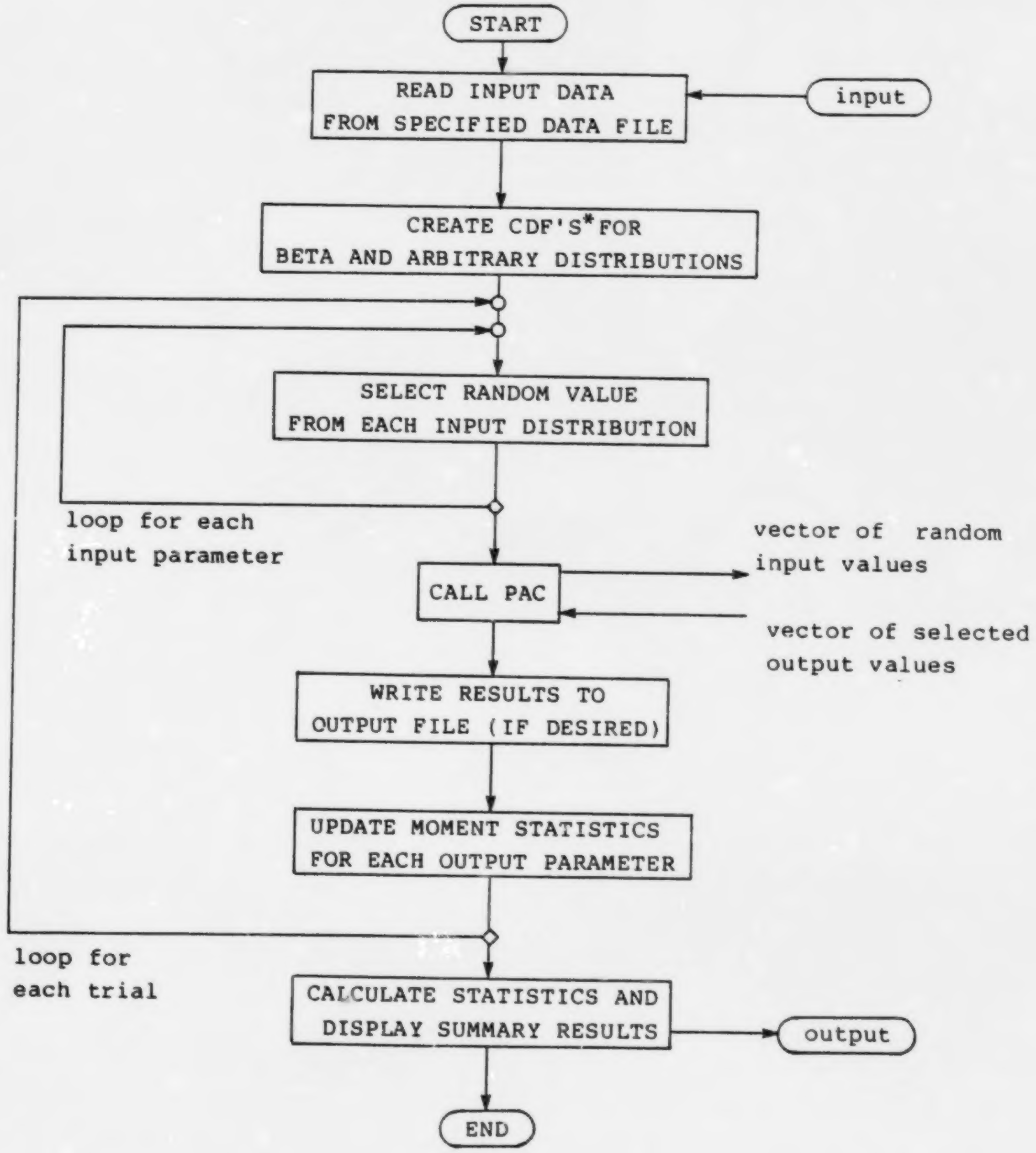
3.2 The Random Simulation Control Module

The probabilistic analysis is driven by a FORTRAN-77 computer program developed by SENES Consultants Limited (SENES 1984). This random simulation program (RANSIM) is a variation on the OPANAL program originally developed by Dr. P. M. Reilly (Reilly and Johri 1969). RANSIM has the capability to draw random samples from a variety of probability distributions, route these values to user-specified simulation models, and perform simple statistical analysis of selected results.

The internal structure of RANSIM is illustrated schematically in Figure 3.2. It is designed to convert numeric descriptors of many types of distributions into a series of randomly-sampled values available for transfer to the simulation models. For example, a normal distribution would be characterized by its mean

FIGURE 3.2

CONCEPTUAL FLOW DIAGRAM OF RANSIM


START
READ INPUT DATA FROM SPECIFIED DATA FILE
input
CREATE CDF'S* FOR BETA AND ARBITRARY DISTRIBUTIONS
SELECT RANDOM VALUE FROM EACH INPUT DISTRIBUTION
loop for each input parameter
CALL PAC
vector of random input values
vector of selected output values
WRITE RESULTS TO OUTPUT FILE (IF DESIRED)
UPDATE MOMENT STATISTICS FOR EACH OUTPUT PARAMETER
loop for each trial
CALCULATE STATISTICS AND DISPLAY SUMMARY RESULTS
output
END

* CUMULATIVE DISTRIBUTION FUNCTION

and standard deviation, while a uniform distribution would be described fully by specifying its maximum and minimum values. The distributions currently available in RANSIM include:

- constant value
- normal
- lognormal
- uniform
- log-uniform
- triangular
- beta
- arbitrary
- correlated

The beta and arbitrary distribution capabilities allow almost an infinite range of distribution shapes to be specified.

Since one often desires immediate feedback from a simulation, RANSIM has the capacity to calculate rapidly a variety of summary statistics on either user-specified input parameters or calculated parameters. Summary statistics available include:

- arithmetic mean
- arithmetic standard deviation
- geometric mean
- geometric standard deviation
- skewness
- kurtosis
- maximum value
- minimum value

These are all calculated using recursive ("on-the-fly") methods and thus can be determined regardless of whether one wishes to save values from all trials or not. The algorithms are described by Spicer (1972). They are most beneficial when the amount of information generated is large and which, if stored, would result in disk storage problems. If one elects to save the results from all trials, they are stored in an output file in a form suitable for subsequent analysis or graphical display by an appropriate computer package (e.g. SAS, SPSS, BMDP, etc.).

3.3 Component Modules and Parameter Allocation and Control

For ease of implementation of a probabilistic analysis, the most important conceptual substructure is the Parameter Allocation and Control (PAC) module. It acts as an interface between the very general RANSIM module and the problem specific simulation modules. While some of its features will be constant over most applications, others will change with each specific problem. PAC is illustrated schematically in Figure 3.3.

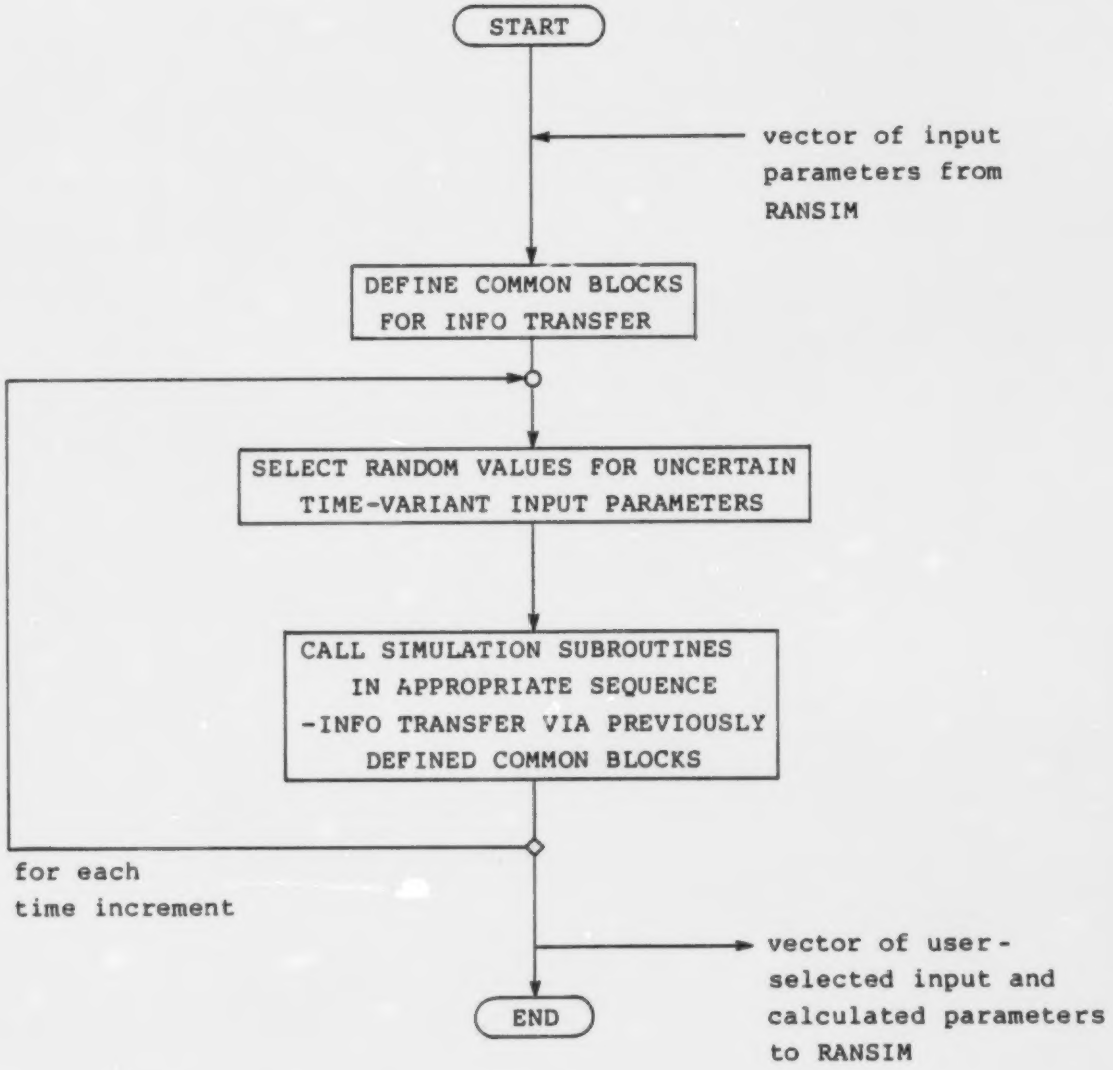
As indicated, PAC accepts a vector of input parameter values from RANSIM and makes them available to the appropriate simulation modules. The PAC module is constructed to access the component modules in an appropriate sequence and make intermediate results available wherever necessary. Finally, it ensures that the appropriate calculated results (and any input parameters of interest) are transferred back to RANSIM for analysis and (possibly) disk storage.

It should also be noted from Figure 3.3 that certain aspects of time dependency are also managed by the PAC module. Since it is often necessary or desirable to perform simulations over varying time periods, this feature is critical in a viable probabilistic analysis program.

The component (simulation) modules are simply computer implementation (in FORTRAN-77) of the mathematical models

FIGURE 3·3

CONCEPTUAL FLOW DIAGRAM OF PAC


START
vector of input parameters from RANSIM
DEFINE COMMON BLOCKS FOR INFO TRANSFER
SELECT RANDOM VALUES FOR UNCERTAIN TIME-VARIANT INPUT PARAMETERS
CALL SIMULATION SUBROUTINES IN APPROPRIATE SEQUENCE -INFO TRANSFER VIA PREVIOUSLY DEFINED COMMON BLOCKS
for each time increment
vector of user-selected input and calculated parameters to RANSIM
END

describing a particular problem. Their interface with the PAC and RANSIM modules is quite straightforward (see Volume 3).

3.4 The Quality Assurance Program

An ongoing quality assurance program was carried out during the Phase II investigations. The overall approach and many of the specific techniques employed were drawn from the Phase I study (SENES 1984). The background theory and current concepts in software quality assurance can be found in an appendix to that report.

This section provides an overview of the quality assurance (QA) program employed during the development of the system model just described. Some control module development was necessary as was extensive development of component models. The software development process involved in sequence, generation of software requirements, software specifications, detailed design and coding of each control and component module.

The QA strategy consisted of many activities aand was integrated into all phases of the software development process. Formal and informal peer reviews (walkthroughs) were employed during the early phases of software development. Once the modules were coded, the QA emphasis shifted to program testing. Highlights of ongoing QA program can be discussed in three principal areas:

1) the Random Simulation Tool
2) the Component Models
3) the System Model.

Further details of each are provided in the following sections.

3.4.1 Tests of the Random Simulation Tool RANSIM

The most general, single objective of RANSIM is to randomly sample values from a variety of user-specified probability distributions. At the heart of this module is the random number generator and the routines specified for the different probability distributions. As part of the overall QA program, the random number generator and distribution sampling routines were subjected to a series of tests.

Random Number Generator

The random number generator UNIRAN, described by Marse and Roberts (1983), was incorporated into RANSIM in Phase II. The program generates uniformly distributed random numbers in the interval 0 to 1 and is claimed by the authors to have excellent global statistical properties and to be highly portable from one computer system to another. The program was subjected to a set of tests however, to verify these claims.

In the first test, the portability of the generator was examined by comparing files of 5000 random numbers produced, using the same seed, on two different computer systems, the inhouse microcomputer facilities at SENES and the mainframe system at the University of Waterloo. The numbers were found to be identical, thus providing support to the authors claim (Marse and Roberts 1983) that the program is portable.

To assess the statistical properties of the random number generator, a second set of 5000 numbers was produced using a different seed. The two sets of numbers were subjected to various statistical tests based on the null hypothesis that the two sets of 5000 numbers consist entirely of independent observations on the uniform (rectangular) distribution covering the interval 0 to 1.

In the first test the mean values of the two sets (0.5000798 and 0.5034284) were considered to be observations on a normally distributed random variable with mean 0.5 and variance 1/(12x5000). Evaluation of the probabilities of obtaining the mean values given for the two sets of 5000 random numbers indicated that they were well within the expected range. This simple test provided initial confirmation that the distribution of the numbers was correct.

As a second test autocorrelation values were calculated. Examination of 20 autocorrelations on each set of random numbers showed that they were within mathematical expectations. Only one autocorrelation in each case was greater than the 95% confidence interval. Hence, it was concluded that none of the autocorrelations were different from zero. The examination of the autocorrelations was extended to an evaluation of the sum of the squares of the autocorrelations. Again, the probabilities of the calculated values being exceeded were found to be within expectations (i.e. not unusual). It was concluded therefore, that the autocorrelations were of a magnitude to be expected from a set of independent random numbers.

The third test applied to the random number generator examined the shape of the frequency distributions constructed from the two sets of 5000 numbers. Application of the chi-squared test, the most generally accepted test of the overall frequency distribution, showed that there was no reason to suspect that the random numbers were not uniformly distributed over the interval 0 to 1.

In the final test applied to the random number sets, the randomness of extreme values was assessed. The 20 smallest and

20 largest numbers from each set were evaluated by a probability integral transformation. Visual inspection of the transformed values suggested they were independent and uniformly distributed as postulated by the null hypothesis. To confirm this observation, the data were subjected to a Kolmogorov-Smirnov test which examines the difference in probability between the postulated cumulative probability function and the sample cumulative probability function. The probability values calculated for the 20 smallest and 20 largest values from both sets of random numbers were all found to be reasonable. Thus it was concluded that the random number generator produces extreme values which have the expected distribution in sets of 5000 random numbers.

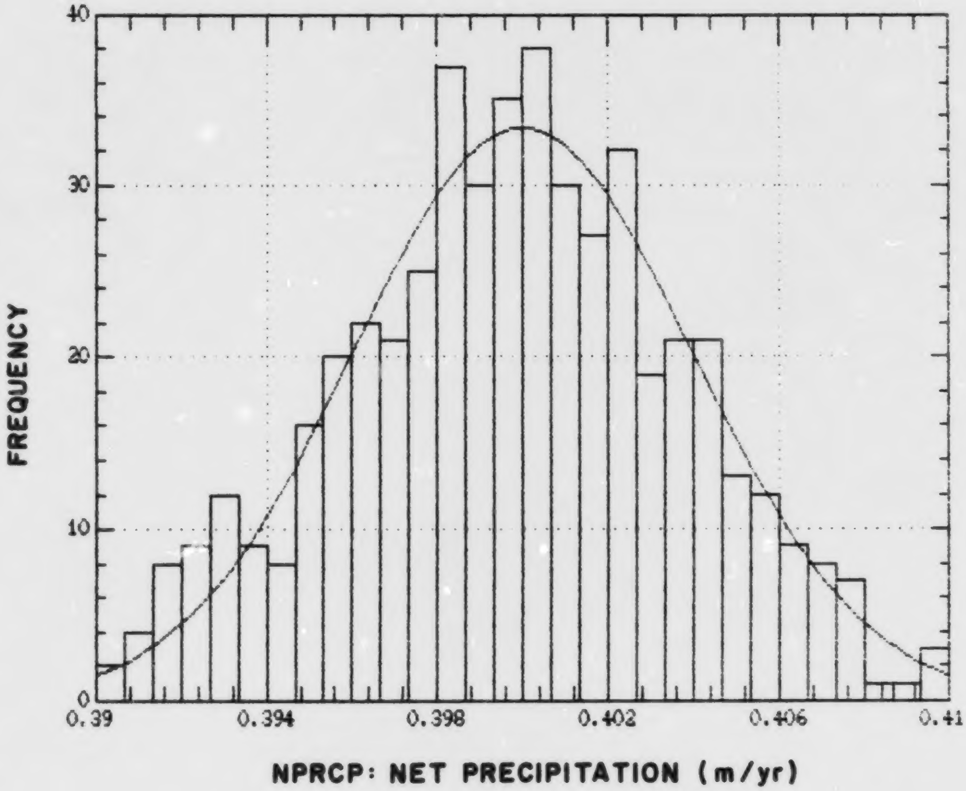
In summary, the tests carried out in this study were found to support the claim of Marse and Roberts (1983) that the UNIRAN generator has excellent global statistical properties. As pointed out by the authors, there may be certain seeds which produce undesirable properties in the random numbers. This is in fact a necessary consequence of the property of randomness of any random number generator that some such sets must exist.

Probability Distribution Routines

Ultimately the test of a random simulation tool is its ability to produce random samples that mimic the attributes of the specified input probability distributions. Random samples of 500 data points were drawn using RANSIM from the normal, lognormal, uniform and triangular distributions for investigation. Frequency histograms were then constructed and are plotted on Figures 3.4 to 3.7 respectively. Also included on the figures for comparison purposes are the shapes of the specified input probability distribution functions. Summary statistics are presented on the bottom of the figures and include the specified

FIGURE 3·4

PLOT OF FREQUENCY HISTOGRAM (500 SAMPLES)
AND PROBABILITY DISTRIBUTION (SMOOTH CURVE)
FOR NORMAL DISTRIBUTION



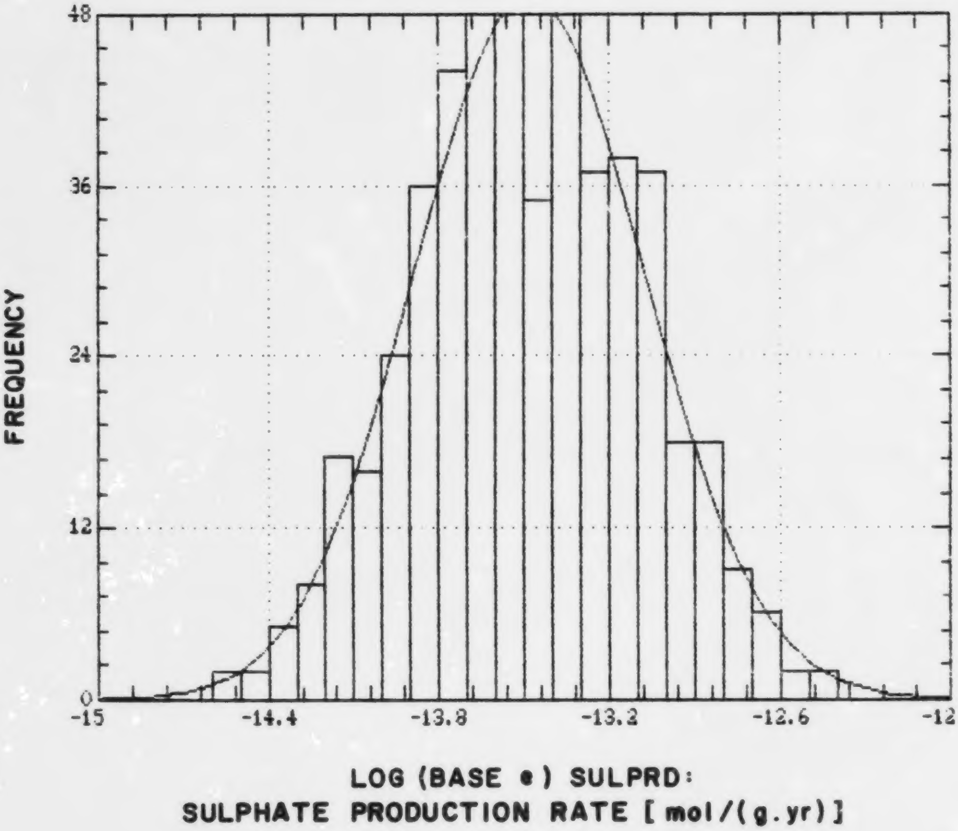
SUMMARY STATISTICS:

	SPECIFIED INPUT VALUES	CALCULATED SAMPLE VALUES
MEAN	0·400	0·400
STANDARD DEVIATION	4×10^{-3}	$3{\cdot}9 \times 10^{-3}$

CHI^2 GOODNESS-OF-FIT = 19·962 WITH 25 DEGREES OF FREEDOM

FIGURE 3·5

PLOT OF FREQUENCY HISTOGRAM (500 SAMPLES)
AND PROBABILITY DISTRIBUTION (SMOOTH CURVE)
FOR LOGNORMAL DISTRIBUTION



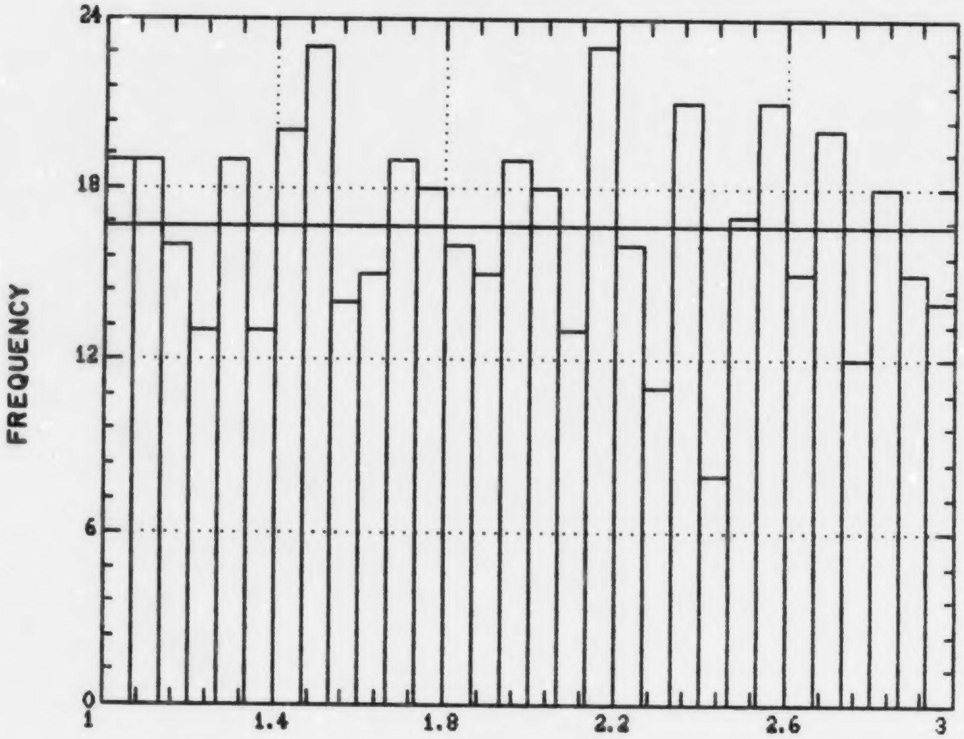
SUMMARY STATISTICS:

	SPECIFIED INPUT VALUES	CALCULATED SAMPLE VALUES
GEOMETRIC MEAN	$1 \cdot 4 \times 10^{-6}$	$1 \cdot 4 \times 10^{-6}$
GEOMETRIC STANDARD DEVIATION	1·5	1·5

CHI^2 GOODNESS - OF - FIT = 13·991 WITH 18 DEGREES OF FREEDOM

FIGURE 3·6

PLOT OF FREQUENCY HISTOGRAM (500 SAMPLES) AND PROBABILITY DISTRIBUTION (SMOOTH CURVE) FOR UNIFORM DISTRIBUTION



DISTCO: RADIUM SOLID TO LIQUID DISTRIBUTION COEFFICIENT (UNSATURATED ZONE)

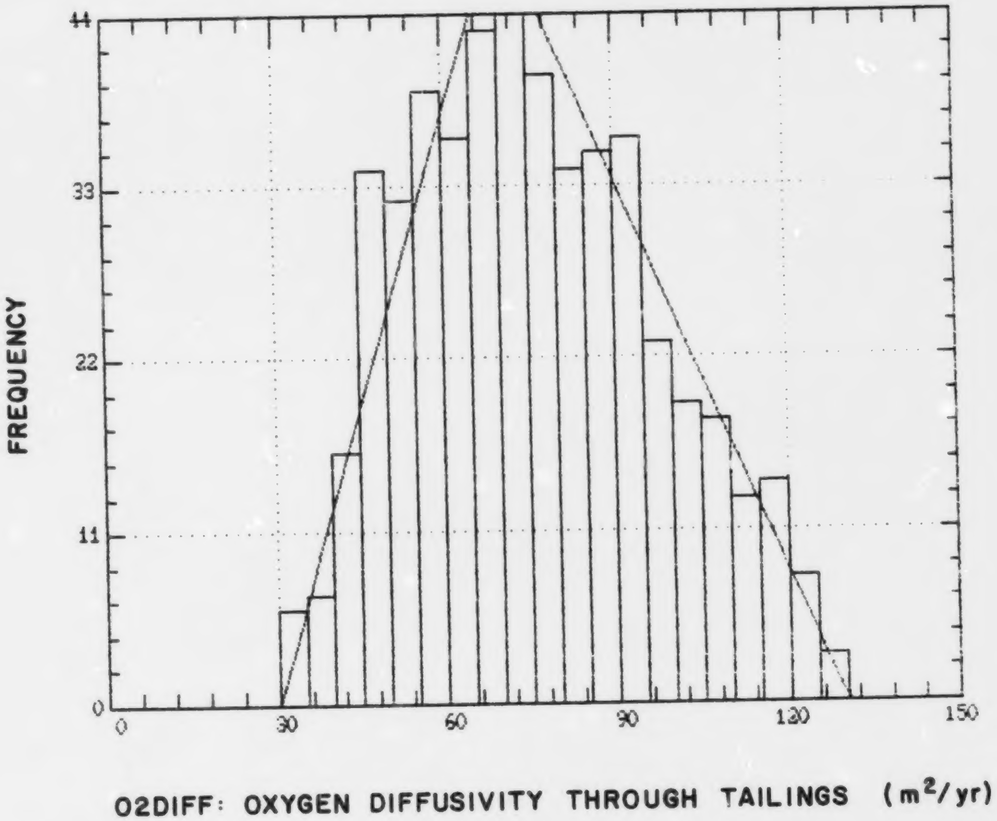
SUMMARY STATISTICS:

	SPECIFIED INPUT VALUES	CALCULATED SAMPLE VALUES
MINIMUM	1	1·01
MAXIMUM	3	2·99

CHI^2 GOODNESS-OF-FIT = 22·12 WITH 29 DEGREES OF FREEDOM

FIGURE 3·7

PLOT OF FREQUENCY HISTOGRAM (500 SAMPLES)
AND PROBABILITY DISTRIBUTION (SMOOTH CURVE)
FOR TRIANGULAR DISTRIBUTION



SUMMARY STATISTICS:

	SPECIFIED INPUT VALUES	CALCULATED SAMPLE VALUES
MINIMUM	30	31·5
MODE	70	-
MAXIMUM	130	129

CHI^{2} GOODNESS-OF-FIT = 14·482 WITH 17 DEGREES OF FREEDOM attributes of the input functions and the calculated statistics for the histograms. The chi-squared goodness-of-fit statistic is also presented on each figure.

Visual inspection of the figures suggests that the shape of the respective distributions are adequately reflected by the histograms in each case. Similarly, the summary statistics show close agreement between the specified attributes of the input functions and the calculated attributes of the computer-generated histograms. For example, the mean values and standard deviations given on Figure 3.4 for the normal distribution are quite comparable.

The chi-squared test, as previously noted, is the most generally accepted test of the discrepancy existing between observed and expected frequencies. The computed chi-squared values are all very ordinary (i.e. the probabilities of obtaining larger values of chi-squared fall in the range from 50 to 75%). Hence, it could be reasonably concluded that the randomly-generated frequency distributions and the specified input probability distributions are the same in all cases.

3.4.2 Component Model QA

The component model QA program involved ensuring that the mathematical formulation and the coding of the individual component models was correct. As initial steps, the specifications, conceptual design and mathematical formulations of the component models were subjected to walkthroughs and peer review. A number of problems were thus identified and corrected before coding began.

Once verified and documented, the component models were coded, and tested individually and in small groups. Initial tests took the form of single deterministic evaluations, using nominal input values. The results, where feasible, were hand-checked to ensure computational correctness. Subsequently, the individual component models were evaluated probabilistically, using the full range of input values, to ensure that the range of output values generated were reasonable. The source term model, being the most complex, was tested more extensively than the other component models. Its time- dependent behaviour was examined using the results of both single deterministic evaluations and small scale probabilistic trials. All of the results were checked for reasonableness by those responsible for the conceptual development of each component model.

3.4.3 System Model QA

The system level QA program involved ensuring the proper integration of the control modules and submodels. The order of integration of the submodels into the probabilistic system model was checked first to ensure that they were linked in a logical sequence. Once integrated, all aspects of the system model were fully tested. Individuals responsible for the development of particular component models were given copies of the results generated by the code to check that the results following integration remained reasonable and correct.

The system model was first run deterministically with all parameters held constant at their nominal values. Those responsible for the development of each model were given a copy of the input parameter specifications, the model coding and a detailed printout of results to check. A number of errors were corrected and the systems model was next run deterministically for two identical trials for two hundred years (i.e. identical input values were specified for both trials). The inputs, code and output were again checked by the formulators. In this case the time performance of the system was checked. As well, the two deterministic trials with identical input parameters were checked to ensure that they yielded identical results (to ensure that the code was free of undetected bugs).

As the QA program continued, the system model was run probabilistically for two trials. Inputs for each trial were checked to ensure that they were being drawn from the correct distributions. A series of small runs of 20 trials were done and several significant output parameters were examined for agreement with the state of knowledge about the physical system. Discrepancies between the two were resolved. This process continued until all parties were satisfied that the models were functioning properly.

As a further QA measure, the system model was again run deterministically; this time using extreme high or low values for all input parameters, to examine the system response to extreme values. Again, any problems indicated were corrected. Finally, the system model was run for many trials (500) and many years (1000). This run was performed to ensure that the simulations did not fail (i.e. blow up) as processes evolved and changed with time.

In general it was difficult to apply rote, mechanical techniques to QA. The QA strategy was found to evolve continually in response to changing situations and needs. During the Phase II investigations, simulations (with up to 500 trials per simulation) were carried out successfully for time periods of up to 1000 years. This practical experience, combined with the QA measures applied gives us confidence that the sytem model will function as specified for a wide range of conditions. Respecification of input parameter distributions within reasonable bounds should not affect the model. It should be noted, however, that in this type of analysis one can occasionally expect to encounter unusual combinations of parameters which circumvent the checks and balances currently built into the system code. Probabilistic analysis cannot be done by rote; the results of each analysis must be carefully examined for "reasonableness". Additional documentation of all components of the system can be found in Volume 3.

4.0 PHASE II PROBABILISTIC ASSESSMENT RESULTS

4.1 Assessment Methodology

The conceptual structure of the uranium tailings assessment program (UTAP) simulation models is shown schematically in Figure 4.1. As mentioned previously, four different disposal scenarios were modelled using this approach. With the exception of the source model, the basic structure of the component models and their interrelations are unchanged from scenario to scenario. The difference among the scenarios lies within the source term (shown as the area enclosed by the dotted line on Figure 4.1). The tailings site, the pathways of exposure and the reference receptor are as described in Section 2. The computer implementation of the system model was outlined in Section 3 and is described in detail in Volume 3. The characteristics of the component models shown on Figure 4.1 were outlined in Section 2 and are described in detail in the Appendices to Volume 3.

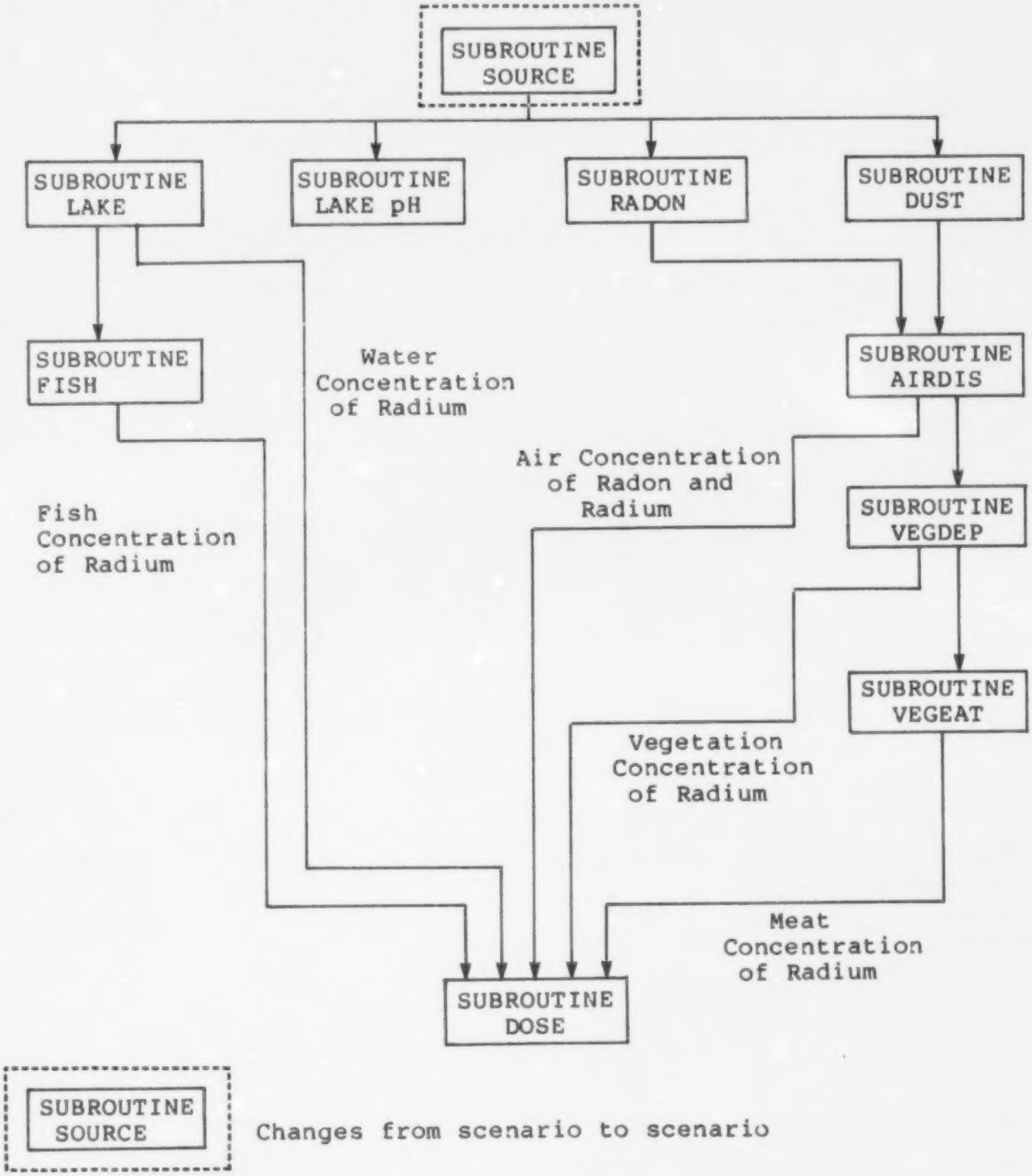
4.1.1 Number of Trials

The number of trials required for a particular simulation depends on the information desired regarding the calculated variable(s). If only an estimate of the mean and its associated standard error is desired, 100 trials is usually more than sufficient. If estimation of the frequency distribution is desired, many more trials will be required.

Except where stated otherwise, the basic statistics and histograms for the analyses described in this section were generated from 500 trials. On the other hand, results from 100 trials were used for regression analysis, in order to provide a data set of manageable size.

FIGURE 4·1

CONCEPTUAL STRUCTURE OF COMPONENT MODEL INTEGRATION


SUBROUTINE SOURCE
SUBROUTINE LAKE
SUBROUTINE LAKE pH
SUBROUTINE RADON
SUBROUTINE DUST
SUBROUTINE FISH
Water Concentration of Radium
SUBROUTINE AIRDIS
Air Concentration of Radon and Radium
Fish Concentration of Radium
SUBROUTINE VEGDEP
SUBROUTINE VEGEAT
Vegetation Concentration of Radium
Meat Concentration of Radium
SUBROUTINE DOSE
SUBROUTINE SOURCE
Changes from scenario to scenario

4.1.2 Implementation of Time Dependency

The program implemented in Phase II to model the system shown on Figure 4.1 is complex and includes several iterative solutions. Consequently, computer run-time (and storage) requirements are fairly large. Since most calculations are performed at each time step, it was found to be important to choose a time step that minimized run-time, while still adequately simulating the time behaviour of the system. It was judged that a time step of 10 years was suitable.

A 1000-year simulation of the base scenario, with 500 trials, demonstrated that beyond about 200 years, the total dose from all pathways was monotonically decreasing. This observation led us to choose a basic simulation period of 500 years thus capturing the essential behaviour of the tailings system while maintaining manageable computer storage and run time. In the simulations discussed in this chapter, parameters which vary from year to year were selected at the start of each trial, from the distribution of 500-year mean values and held fixed throughout the trial (for further discussion of this matter see Section 2.6 and the discussion paper by Beals et al, Appendix D, Volume 2).

Time-dependent calculated results were statistically analysed (and stored) for 50-year time steps; that is, at 11 times from 0 to 500 years. This was found to represent adequately the time behaviour of the system.

4.2 Scenario 1: The Base Case

The base case scenario assumes that the reference tailings site has the characteristics outlined in Section 2.1 of this report and that the reference receptor has the characteristics outlined in Section 2.3. The environmental pathways analysed probabilistically and discussed in this section are those shown on Figure 4.1.

Probabilistic analysis is a complex process that generates an enormous amount of data. The following presentation of the base case scenario attempts to distill the relevant information from this process in a format, and at a level of technical rigour consistent with the goals of an environmental assessment.

The focus of the discussion is on the time behaviour of the overall system, including the relative contribution of different pathways to total dose. Sensitivity/uncertainty analysis of the parameters and pathways contributing to total dose is provided as is a similar analysis of parameters contributing to selected pathways models. A similar analysis is provided for the pH model. The base case is also used to illustrate some interesting features of the temporal and spatial behaviour of the system.

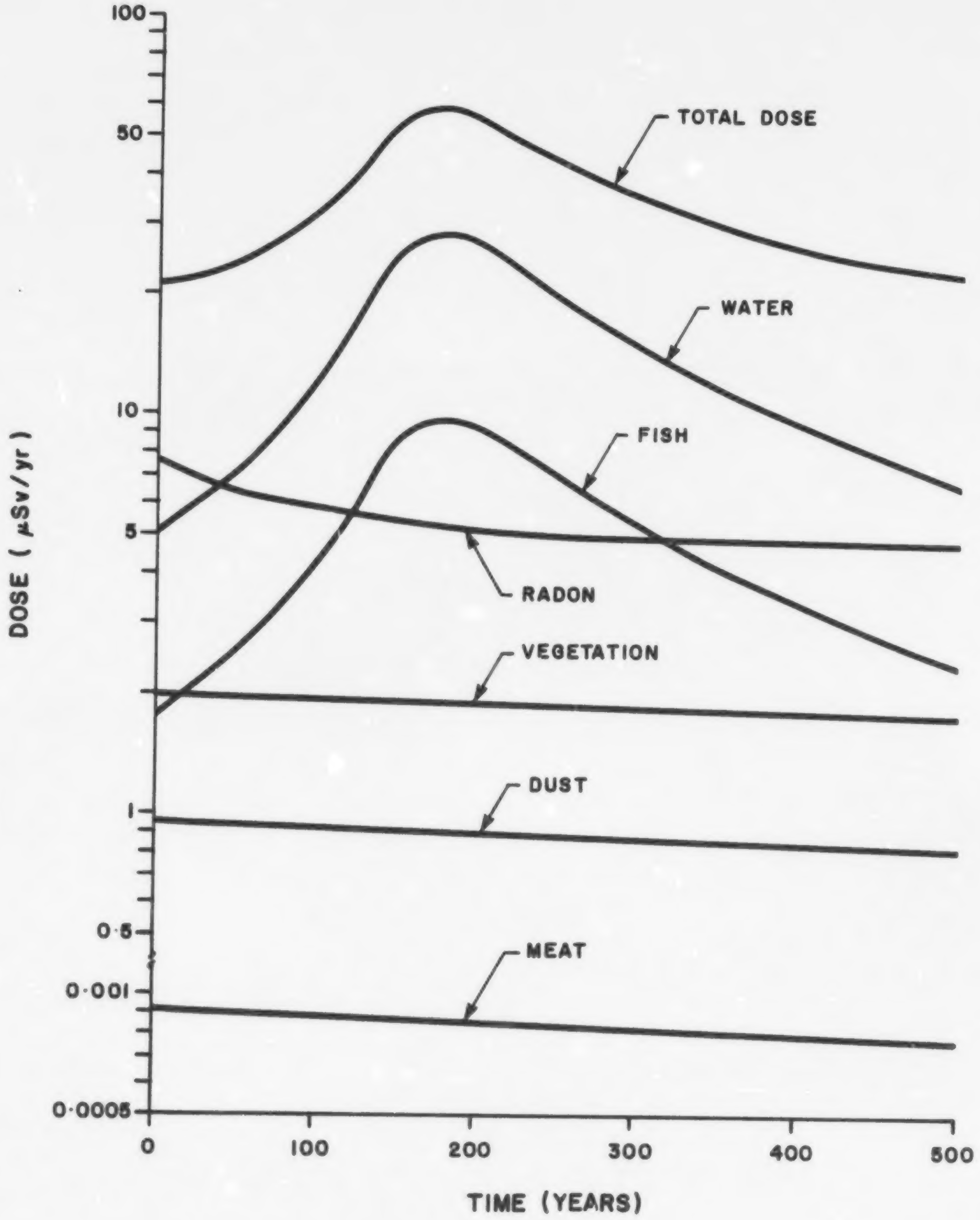
4.2.1 Overall Evaluation

The time behaviour of the geometric mean of total dose and the dose from the six contributing exposure pathways are shown on Figure 4.2. Since the dose histograms tended toward lognormality, the geometric mean is a good indicator of the central tendency of the data. The dose from water-related pathways increases for approximately the first 180 years and decreases thereafter. The air-related pathways decrease monotonically with time.

At the start of the simulation (T=0), radon inhalation is the largest contributor to total dose. Within the initial 150 years of the simulation, radon is surpassed by the dose contributions from consumption of both water and fish. Toward the end of the simulation (T≥300 years) the contribution of the fish pathway has decreased and radon has become the second most important pathway.

FIGURE 4·2

DOSE vs. TIME FOR SCENARIO I
GEOMETRIC MEAN OF 500 TRIALS



In the long term the radon pathway will again contribute the greatest portion of the total dose.

The dose from consumption of contaminated meat is by far the smallest contributor. Indeed, it is two orders of magnitude smaller than the next highest dose contributor (the dust inhalation pathway) and is not a significant contributor to total dose at any time during the simulation period.

The pattern of time behaviour is explained by the gradual dissolution of gypsum in the unsaturated zone of the tailings source during the first one hundred to two hundred years. As discussed in Section 2 and Appendix A of Volume 3, gypsum contains that portion of the radium which is most soluble. Since the gypsum/radium coprecipitate exists in equilibrium with the tailings porewater, the gradual dissolution of gypsum also results in a gradual release of radium. After the gypsum (and associated radium) contained in the upper unsaturated zone is depleted, the primary source of radium is that resulting from the decay of thorium-230, which is partially soluble in the acidic porewater of oxidizing pyritic tailings. This source of dissolved radium is hypothesized to be much less important than that from gypsum dissolution. The levels of radium in the tailings porewater of the unsaturated and saturated zones gradually decrease with time as the gypsum-related radium is flushed from the system.

Radon, on the other hand, is derived from radium associated with both the host tailings material and gypsum solids in the upper unsaturated layer of the tailings. Thus, as the gypsum-related radium is leached from this layer, the radon-222 release will decrease. After gypsum-related radium is depleted from the unsaturated zone (upper layer) the source of radon-222 is determined by the balance between the radioactive decay of radium-226, the ingrowth of radium-226 from thorium-230 and the gradual leaching of some of the radium-226 which ingrows to the thorium-230 matrix. This occurs as a slowly decreasing function of time.

The pH of the tailings porewater is controlled by the oxidation of pyrite, which occurs relatively rapidly over the first 100 to 200 years. As the pyrite content gradually declines in the unsaturated zone, acid generation is greatly reduced and the pH gradually increases toward neutral conditions.

The behaviour of lake pH with time is affected by the pH of tailings porewater and is shown on Figure 4.3. At the start of the simulation, the mean pH equals about 4.0 and gradually increases with time (over the initial 300 years) to about 7.3. Thereafter it remains at that level. As indicated on the figure, the confidence interval on the mean pH is quite wide over the initial 300 years and narrows considerably thereafter. The high variability in the early years reflects the uncertainty of the acid input to the system due to pyrite oxidation. This oxidation is predicted to be essentially complete in the initial 200 to 300 years. The limited variability in pH beyond year 300 is a reflection of the natural background variation which in undisturbed systems, in spite of low buffering capacity, is quite small.

Figures 4.4, 4.5, and 4.6 are histograms of total dose from all six pathways (at T=0, and T=200 years), dose from the radon pathway and dose from the drinking water pathway respectively. All distributions are skewed, with tails stretching to the right, behaviour suggestive of lognormality. Basic, summary statistics are also shown in these figures.

The geometric mean annual total dose at 200 years is about 2.5

FIGURE 4·3

pH vs. TIME FOR SCENARIO I

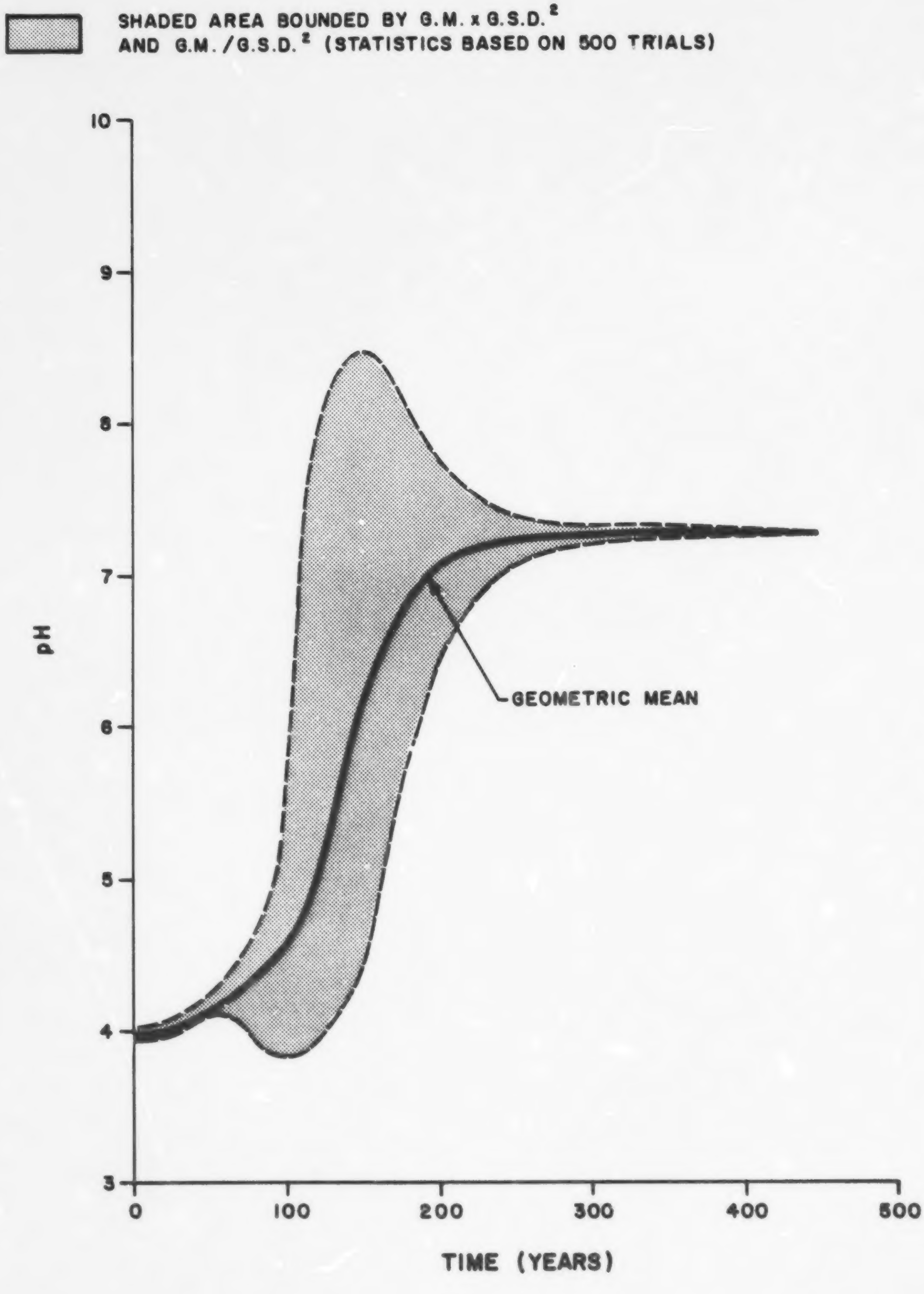
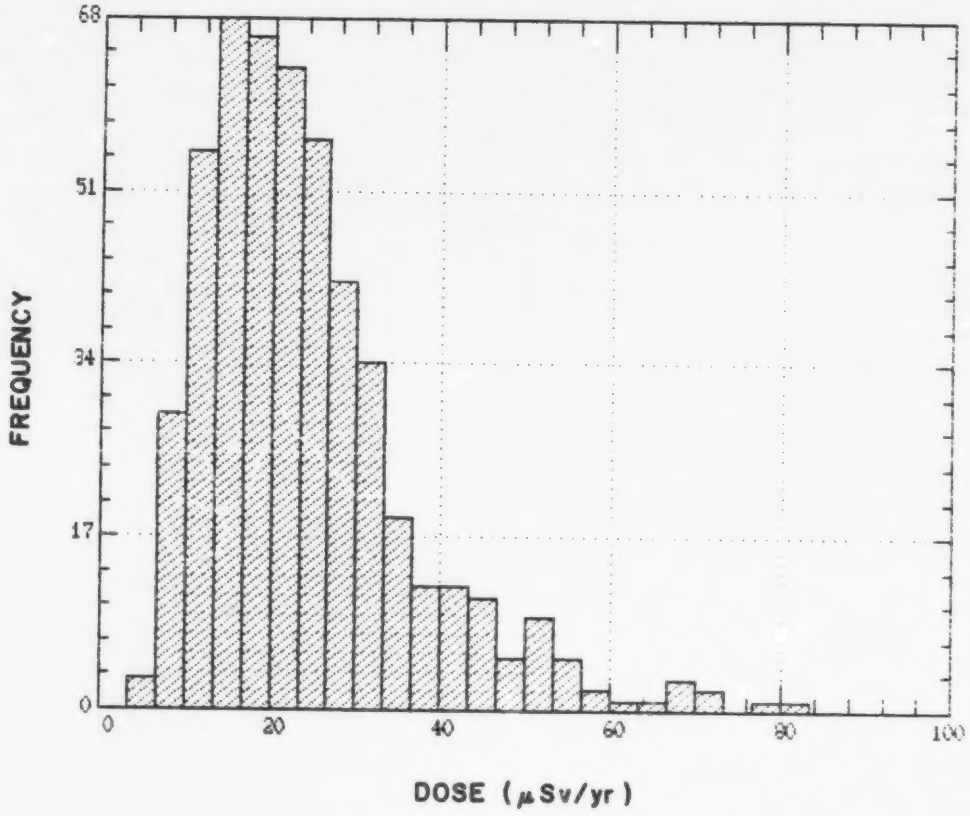


FIGURE 4·4a

HISTOGRAM OF TOTAL DOSE AT T=O FOR SCENARIO I

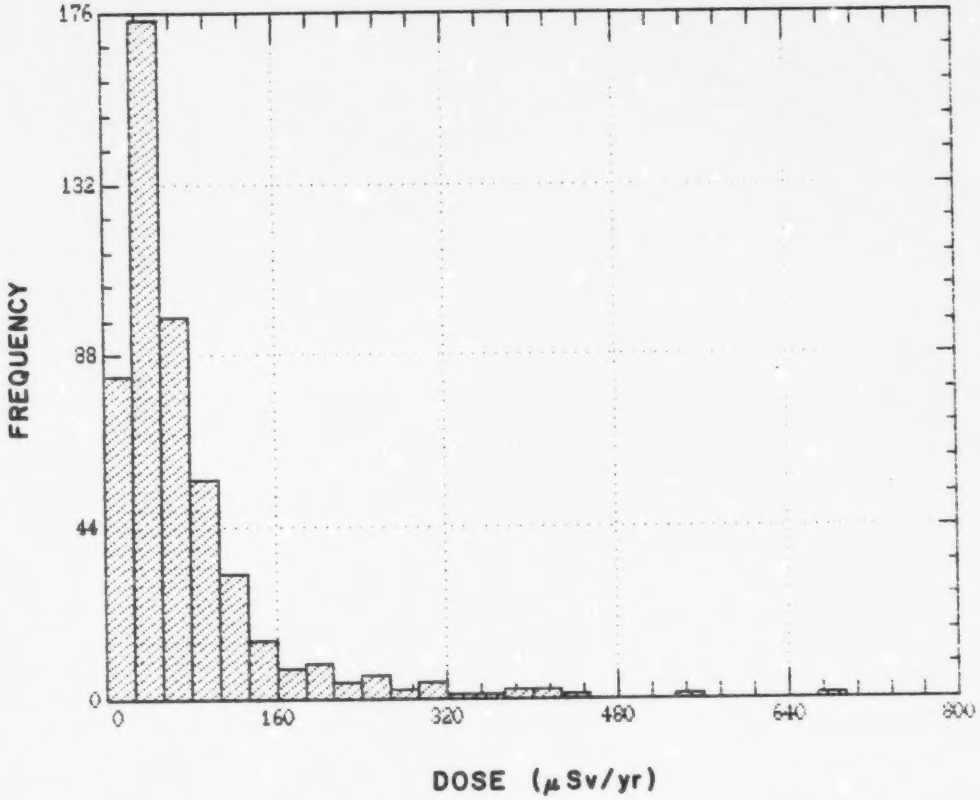


SUMMARY STATISTICS:

MEAN = 24·0
STANDARD DEVIATION = 12·4
GEOMETRIC MEAN = 21·2
GEOMETRIC STANDARD DEVIATION = 1·64
MINIMUM = 4·29
MAXIMUM = 80·9

FIGURE 4·4b

HISTOGRAM OF TOTAL DOSE AT T = 200 FOR SCENARIO I

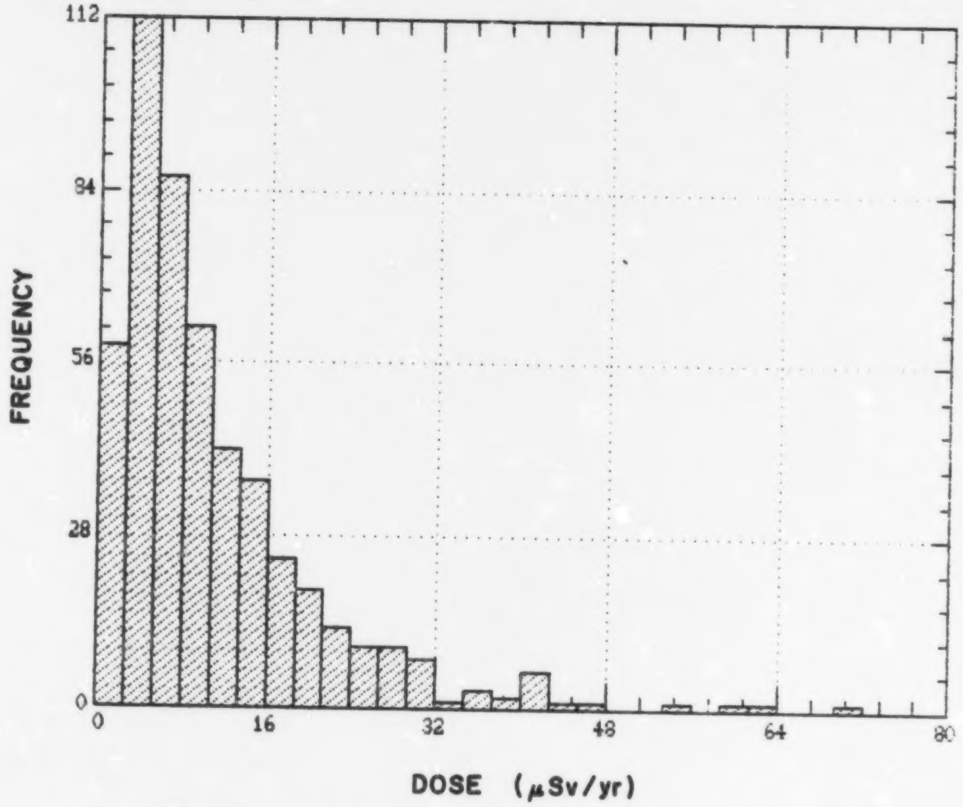


SUMMARY STATISTICS:

MEAN = 75·0
STANDARD DEVIATION = 74·5
GEOMETRIC MEAN = 55·0
GEOMETRIC STANDARD DEVIATION = 2·13
MINIMUM = 5·56
MAXIMUM = 685

FIGURE 4·5

HISTOGRAM OF DOSE FROM RADON INHALATION AT T = 0 FOR SCENARIO I

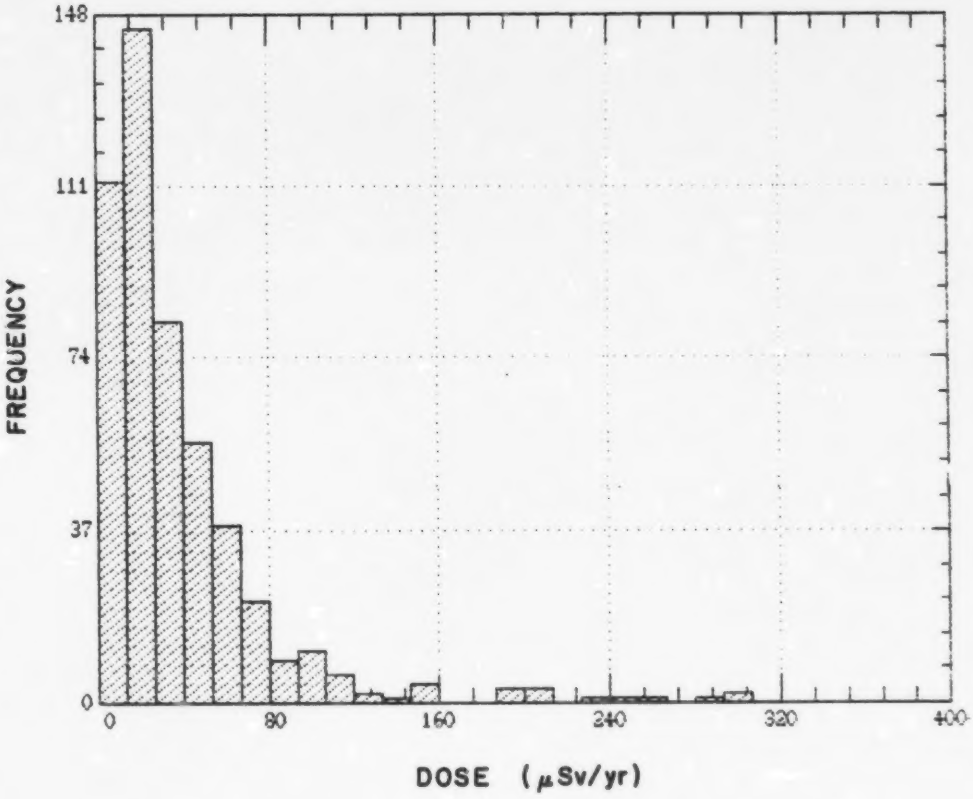


SUMMARY STATISTICS:

MEAN = 10·8
STANDARD DEVIATION = 9·79
GEOMETRIC MEAN = 7·59
GEOMETRIC STANDARD DEVIATION = 2·38
MINIMUM = 0·576
MAXIMUM = 71·3

FIGURE 4·6

HISTOGRAM OF DOSE FROM DRINKING WATER AT T = 200 FOR SCENARIO I



SUMMARY STATISTICS:

MEAN = 38·7
STANDARD DEVIATION = 41·6
GEOMETRIC MEAN = 26·4
GEOMETRIC STANDARD DEVIATION = 2·36
MINIMUM = 2·27
MAXIMUM = 302

times greater than the geometric mean annual total dose at zero years. At zero years the radon pathway contributes about 44% of the total dose whereas the drinking water pathway contributes 29%. At 200-years the drinking water dose contributes about 60% of the total dose while the radon pathway contributes about 12%.

4.2.2 Further Considerations of Temporal and Spatial Dependence

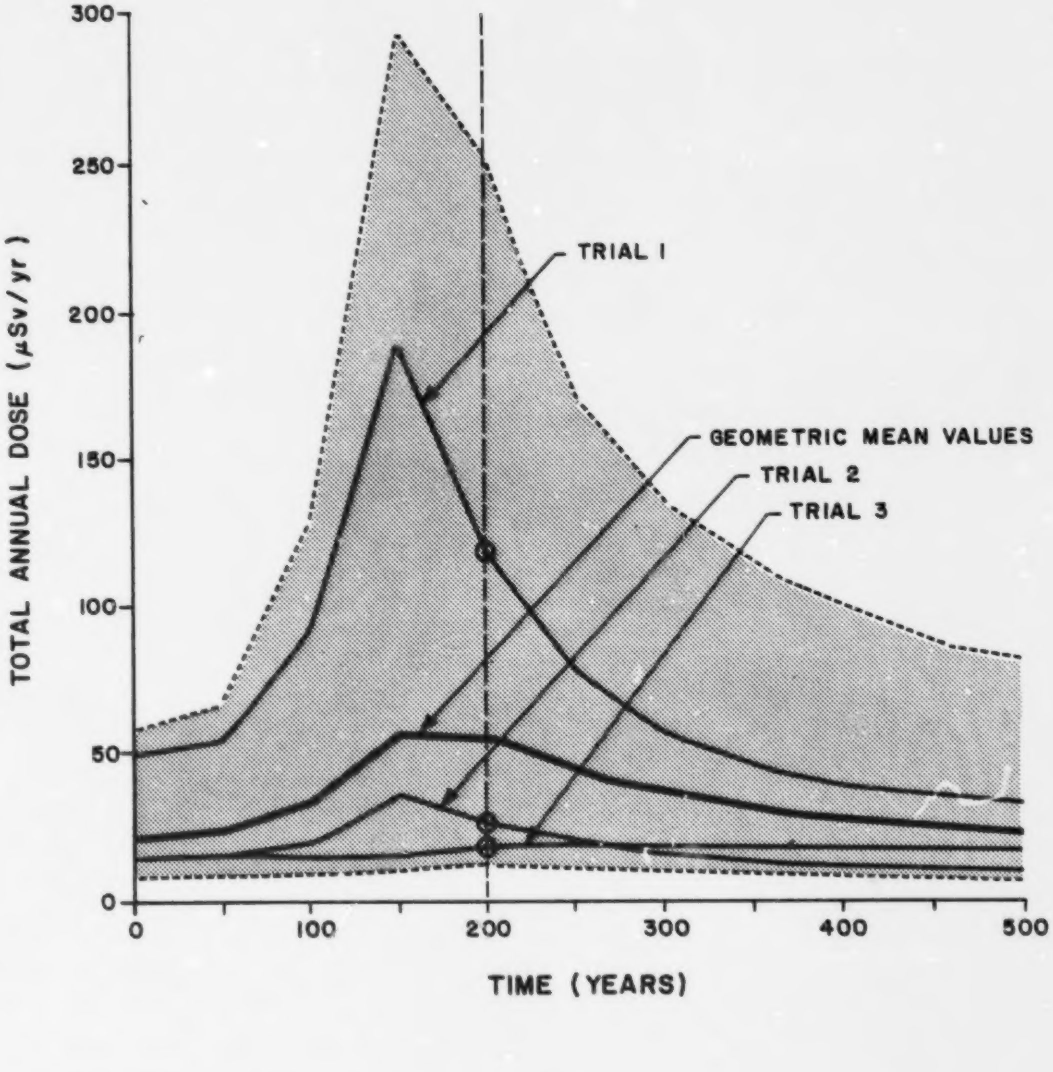
In the introductory sections of this report (eg. Figure 2.2) probabilistic analysis was discussed in terms of repeated evaluations of a simulation model using randomly-sampled values of many input parameters. The results of such analysis were presented in the previous section. Examination of such results (eg. geometric mean value time curves or histograms of particular consequences) obscures some of the finer details of the temporal and spatial dependence of the system. The examples in this section are presented to illustrate these more detailed features.

Figure 4.7 presents both deterministic and probabilistic results as a function of time. The three curves shown on the figure labelled Trial 1, Trial 2 and Trial 3 are the simulation results obtained from three separate computer trials wherein parameter values were randomly selected at T=0 years and the system response was allowed to evolve in time to the end of the simulation period. These curves are, in fact, the first three trials of the 500 trials which were used to construct Figures 4.2 through 4.6. The curves labelled Trial 1,2 and 3 on Figure 4.7 are plots of eleven output data points representing 50 year time increments (i.e. T=0, T=50...T=500 years).

Thus, on Figure 4.7, a line drawn parallel to the vertical axis will intersect each time curve (1 curve per trial) exactly once. Hence, for 500 trials there are 500 time curves and 500 values of total dose at any time. The three small circles shown on the

FIGURE 4·7

TEMPORAL CONSIDERATIONS IN PROBABILISTIC ANALYSIS



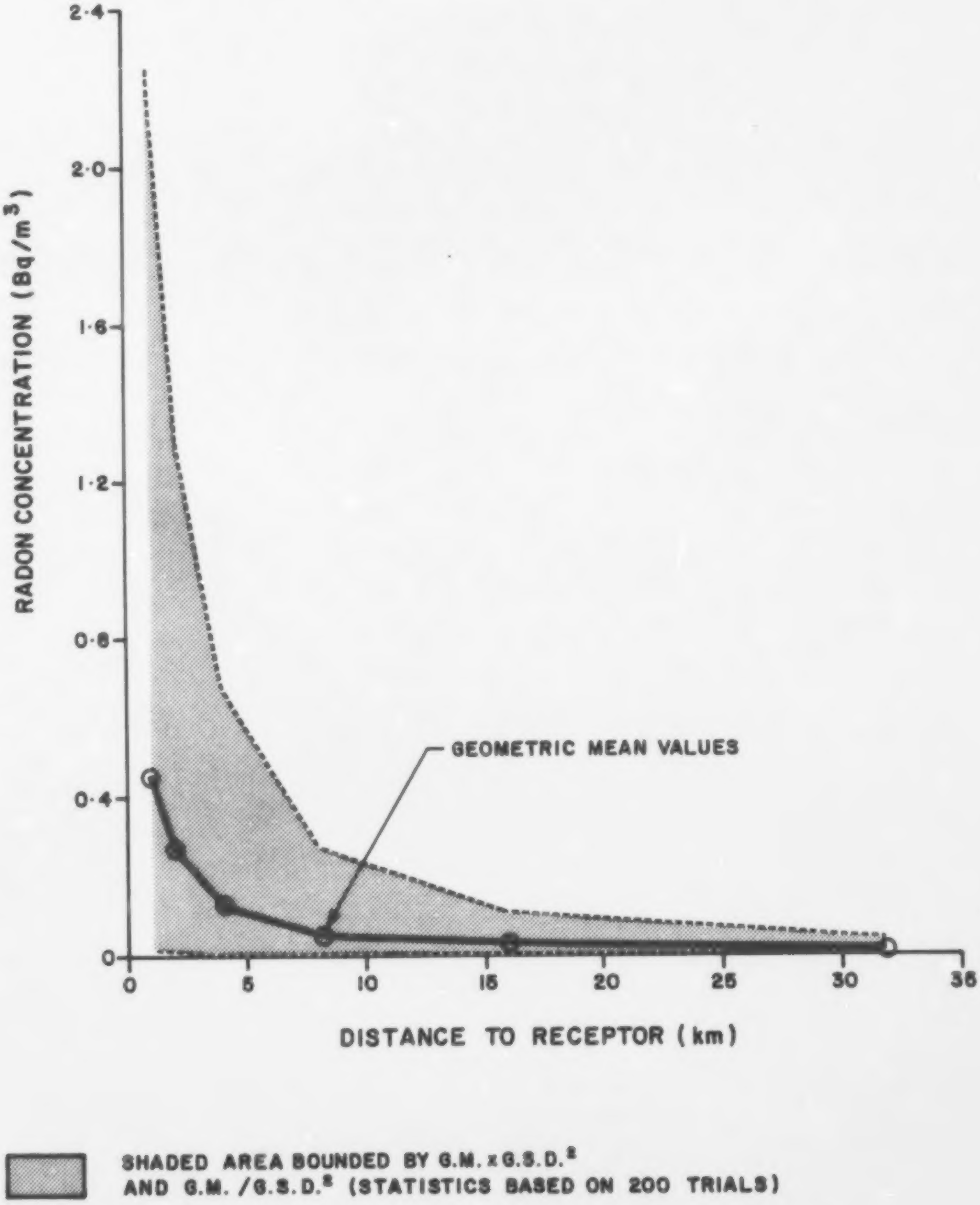
▨ SHADED AREA BOUNDED BY G.M. x G.S.D.^{2} AND G.M./G.S.D.^{2} (STATISTICS BASED ON 500 TRIALS) time curves for trials 1,2,3, lying directly above T=200 years are three of the 500 values used to calculate the summary statistics appropriate for the two-hundredth year. The thicker solid line is a plot of the geometric mean value of 500 such trials, again representing each of the eleven output time steps. The geometric mean was used since the dose distribution tended toward lognormality. The shaded area within the dotted lines is the envelope formed by the 95% probability region. This was determined by multiplying or dividing the geometric mean value by the square of the geometric standard deviation.

In addition to producing results which vary with time, the models also produce results which vary with distance from the tailings. As an example, Figure 4.8 shows how the ambient radon concentration changes with increasing distance from the tailings for the first year of the simulation. The curve would be different at later years. The heavy middle line on the plot joins the geometric mean (of 200 trials) calculated for distances of 1,2,4,8,16 and 32 km from the tailings area. These points are shown as small circles on the figure. The shaded area, defined by the dotted lines, is the envelope formed by the 95% probability region.

The meaning of Figures 4.7 and 4.8 are different. Figure 4.8 is a snapshot at a particular time (T=0 years in this case) of radon concentration versus distance. Figure 4.7 shows the time behaviour of total dose to a particular receptor (as does Figure 4.2). A completely different set of time curves would have to be calculated for a receptor with different characteristics. Together, the curves illustrate many features of probabilistic assessment.

FIGURE 4·8

SPATIAL CONSIDERATIONS
IN PROBABILISTIC ANALYSIS



4.2.3 Scenario 1 Importance Analysis: Total Dose

This section presents an importance analysis of total dose for the base case scenario. As was indicated in Section 2.6, stepwise regression and the sensitivity index were the two importance analysis tools used. Detailed results of stepwise regression and sensitivity index analysis are provided in Appendix A. The reader should refer to the foldout table (Appendix B) at the end of this volume, for parameter names, definitions and input values.

The ten most important parameters at T=0 years and T=200 years from both importance analysis techniques are detailed in Table 4.1 and 4.2 respectively. The parameters are identified in Column 1. Columns 2 and 3 of the table show the ranking according to sensitivity index and stepwise regression respectively for the top ten parameters.

The 4th and 5th columns of Table 4.1 and 4.2 respectively provide a brief description of each parameter and indicate the component models where the parameter first occurs.

The 6th column shows the increase in R^2 of the overall regression attributable to adding the particular variables, during stepwise regression. This is a variance based measure of parameter importance, since the overall R^2 is a measure of the amount of variability in the data (i.e. output values) which is accounted for by the regression. The greater the increase in R^2 when a particular parameter is added to the regression, the more important it is with respect to the output values. As was indicated in Section 2.7.2, the parameter rankings for stepwise regression are derived from their order of inclusion in the stepwise regression and not from the increase in the overall R^2.

Table 4.1

PARAMETER IMPORTANCE RELATIVE TO TOTAL DOSE AT T=0 YEARS FOR SCENARIO 1

(1) PARAMETER *	(2) RANK S.I.	(3) RANK REGRESSION	(4) DESCRIPTION	(5) MODEL	(6) INCREASE IN R^2	(7) S.I. (INFLUENCE)
BF	1	4	water to aquatic produce transfer coefficient: fish flesh	Water to Aquatic Produce	.0910	.9139 (+)
CONCRA	2	2	initial solid phase activity of Ra-226 in tailings	Source	.1710	.6667 (+)
RAINTH	3	8	mole fraction of Ra-226 ingrowth to solid Th-230 matrix which is insoluble	Source	.0138	.5340 (-)
OF	4	5	fraction of year in location	Water Consumption/ Inhalation (radon and particulates)	.0729	.4456 (+)
DISTCO	5	7	radium solid/liquid distribution coefficient (unsaturated zone)	Source	.0231	.4397 (+)
XK1	6	34	rate of radium removal from the lake water column to lake sediment	Lake	.0007	.2948 (-)
EMANAT	7	3	radon emanation coefficient of tailings	Radon Source Term	.1406	.2787 (+)
RHOTAL	8	9	tailings bulk density	Source	.0152	.2765 (+)
RAINGP	9	38	fraction of total Ra-226 which is associated with gypsum	Source	0.0003	.2464 (+)
FISHIN	10	10	average annual intake of fish	Fish Ingestion	.0129	.2014 (+)
MPCH20	12	1	Percent moisture content by weight of tailings	Radon Source Term	.2931	.1656 (-)
F	15	6	equilibrium factor for radon daughters to radon	Inhalation Model (Radon)	.0741	.0959 (+)

* see foldout at end of report for parameter listing

Table 4.2

PARAMETER IMPORTANCE RELATIVE TO TOTAL DOSE AT T=200 YEARS FOR SCENARIO 1

(1) PARAMETER *	(2) S.I.	(3) RANK REGRESSION	(4) DESCRIPTION	(5) MODEL	(6) INCREASE IN R^2	(7) S.I. (INFLUENCE)
BF	1	5	water to aquatic produce transfer coefficient: fish flesh	Water to Aquatic Produce	.0610	.9250 (+)
RAINGP	2	2	fraction of total Ra-226 which is associated with gypsum	Source	.0956	.8111 (+)
DISTCO	3	3	radium solid/liquid distribution coefficient (unsaturated zone)	Source	.0780	.7642 (+)
CONCRA	4	4	initial solid phase activity of Ra-226 in tailings	Source	.0900	.6667 (+)
NPRCP	5	**	net precipitation	Source	---	.5852 (+)
FACTCO	6	10	activity coefficient (for divalent ions)	Source	.0126	.4821 (-)
OF	7	8	fraction of year in location	Inhalation (radon and particulates) Water Consumption	.0210	.4474 (+)
KSUBDR	8	1	Ra liquid-solid transfer coefficient (mixing zone)	Source	.2854	.4443 (+)
SULPRD	9	45	sulphate production rate	Source	.0003	.2719 (+)
RHOTAL	10	7	tailings bulk density	Source	.0253	.2435 (+)
EMANAT	16	9	radon emanation coefficient of tailings	Radon Source Term	.0158	.0826 (+)
MPCH20	19	6	percent moisture content by weight of tailings	Radon Source Term	.0415	.0447 (-)

* see foldout at end of report for parameter listing

** Not included (eliminated) in stepwise regression.

The final column, column 7, shows the sensitivity index and influence. The sensitivity index has been explained previously. The "influence" refers to the direction of response of total dose to a change in magnitude of a particular parameter. Plus (+) means a direct proportionality while minus (-) indicates an inverse proportionality.

With exception of parameters XK1 and RAINGP, the results produced by the two importance analysis methods are similar. Of note is the dominance of the source model parameters. Full printouts of stepwise regression analysis and the sensitivity index for total dose are provided in Appendix A.

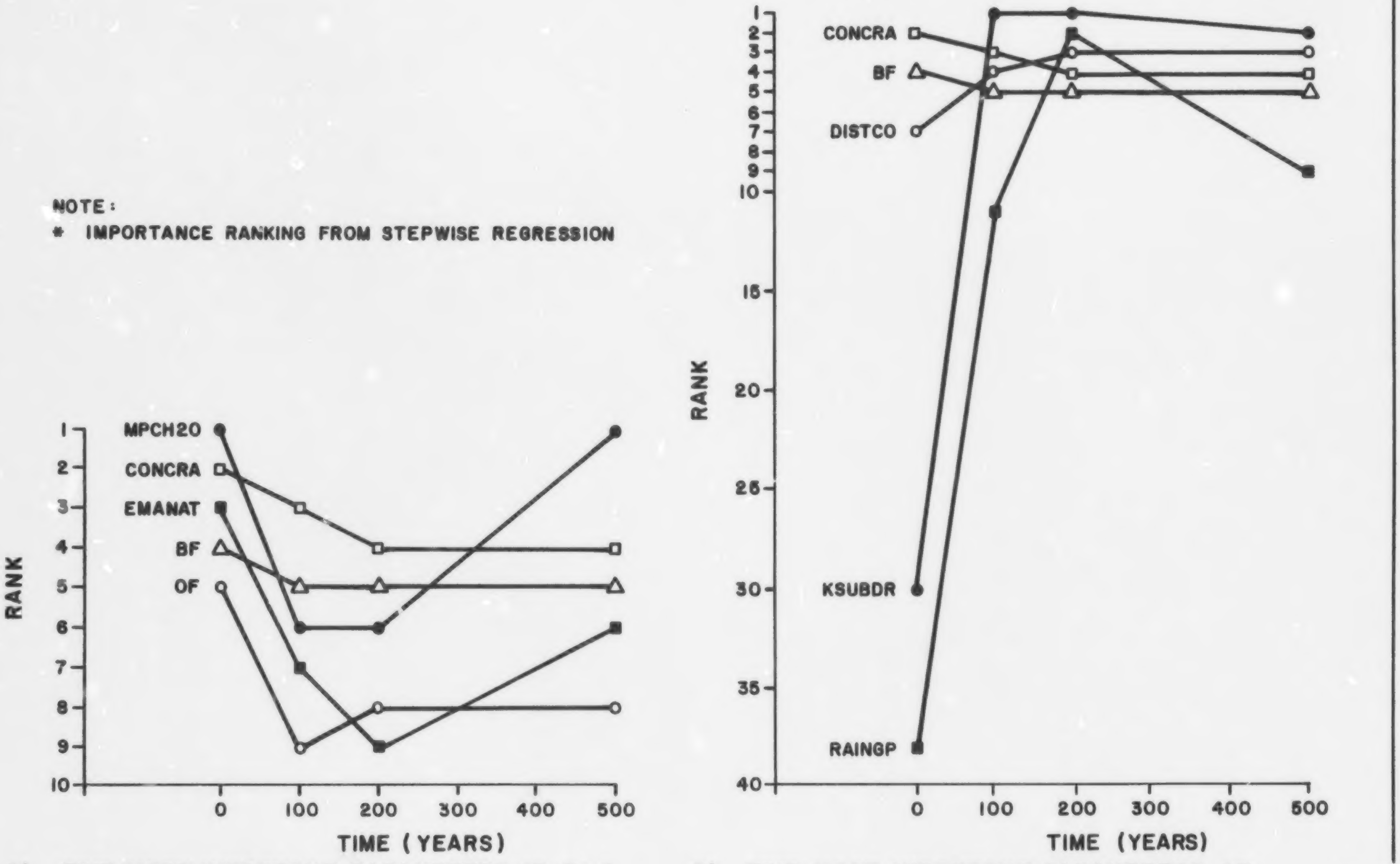
It is of interest to follow the change in parameter importance with time. Figure 4.9(a) shows how the importance of the five top ranked parameters, selected by stepwise regression at T=0 years, changes with time. All of the parameters remain in the "top ten" (as determined by stepwise regression at the indicated timesteps) throughout the full simulation period. At T=0 years the parameters related to the radon pathway (MPCH20, CONCRA and EMANAT) are most important. During the period between years 100 to 200, two parameters related to the water and fish consumption pathways (CONCRA and BF) are of prime importance. This corresponds to the period during which the dynamics of radium dissolution in the tailings are greatest.

Figure 4.9(b) shows how the importance of the five top ranked parameters, selected by stepwise regression at T=200 years, changes with time. The behaviour in this instance is dramatically different from that evidenced in Figure 4.9(a). Only three of the parameters (CONCRA, BF and DISTCO) are found in the top ten over the entire simulation period. Two parameters (KSUBDR and RAINGP), which are of relatively low importance at T=0 years rise to first and second in importance in the period

FIGURE 4·9

CHANGE IN PARAMETER IMPORTANCE RANKING vs. TIME

NOTE:
* IMPORTANCE RANKING FROM STEPWISE REGRESSION



A) FIVE MOST IMPORTANT PARAMETERS AT T=0 YEARS*

B) FIVE MOST IMPORTANT PARAMETERS AT T=200 YEARS* from 100 to 200 years. As noted previously this is the period of maximum biological and chemical activity in the upper tailings layer (arising from bacteria-assisted pyrite oxidation and gypsum dissolution and reprecipitation). Following this period of active chemical behaviour, the parameter for the radium solid-to-liquid transfer coefficient in the mixing zone (KSUBDR) remains highly ranked while the parameter radium-in-gypsum (RAINGP) decreases in importance, as gypsum is depleted from the upper tailings horizon.

It is also of interest to identify the least important parameters. Table 4.3 contains a listing of the ten least important parameters. The parameters were ranked according to stepwise regression and according to sensitivity index at zero and 200 years. The ten least important parameters overall, were then determined according to the average of the aforementioned ranks. These parameters are associated with meat and vegetable consumption and dust inhalation, consistent with the results shown in Figure 4.2.

4.2.4 Importance Analysis: Selected Pathways

The results of a sensitivity index analysis and stepwise regression of the water pathway at T=200 years and the radon inhalation pathway at T=0 years are shown in Tables 4.4 and 4.5 respectively. The information in the tables was derived as follows. First, a sensitivity index analysis was performed on all model parameters with respect to water consumption dose (at T=200 years) and radon inhalation dose (at T=0 years). This procedure not only helped isolate important parameters, but also provided a ready check on parameters not involved in the pathway's calculation. (The parameters not involved have a S.I. of zero.) Stepwise regressions were then performed on the parameters involved in the calculations for the two pathways.

Table 4.3

TEN LEAST-IMPORTANT PARAMETERS FOR SCENARIO 1

PARAMETER	DESCRIPTION	MODEL
VEGLAM(2)	decay constant accounting for weathering loss (pasture grass)	Vegetation Deposition
DUREXP(1)	duration of exposure to deposition (lichens)	Vegetation Deposition
FMEAT	fraction of annual meat intake that is local caribou	Caribou Ingestion
HSRC	height of the source	Air Dispersion
VEGLAM(4)	decay constant accounting for weathering loss (below ground vegetation)	Vegetation Deposition
DUREXP(2)	duration of exposure to deposition (pasture grass)	Vegetation Deposition
FDEPSF(4)	fraction of total deposition retained on surface of vegetation (below ground vegetation)	Vegetation Deposition
FACTB	fraction of activity in below ground vegetation after preparation	Vegetation Dose
FWIND1	percent of time that the wind speed exceeds 5.4 m/s at the mean tailings height	Dust Source
VEGLAM(3)	decay constant accounting for weathering loss (above ground vegetation)	Vegetation Deposition

Table 4.4

PARAMETER IMPORTANCE RELATIVE TO DOSE VIA WATER CONSUMPTION PATHWAY AT T=200 YEARS FOR SCENARIO 1

	RANK				
PARAMETER*	S.I.	REGRESSION	MODEL	INCREASE IN R^2	S.I. (INFLUENCE)
RAINGP	1	2	Source	0.1832	.8290 (+)
DISTCO	2	3	Source	0.1217	.7737 (+)
CONCRA	3	4	Source	0.0628	.6667 (+)
NPRCP	4	19	Source/lake	0.0061	.5977 (+)
OF	5	5	Water Consumption	0.0274	.5556 (+)
FACTCO	6	7	Source	0.0134	.4955 (-)
KSUBDR	7	1	Source	0.3134	.4914 (+)
SULPRD	8	11	Source	0.0023	.2815 (+)
RHOTAL	9	6	Source	0.0191	.2462 (+)
KPRIME	10	15	Source	0.0001	.2422 (-)
RAINTH	11	8	Source	0.0086	.2149 (-)
BETA	12	16	Lake	0.0001	.1351 (+)
XK3	13	10	Lake	0.0039	.0897 (-)
O2DIFF	14	12	Source	0.0019	.0817 (-)
XK1	15	not entered	Lake	not entered	.0688 (-)
THFRAC	16	14	Source	0.001	.0282 (+)
THDISS	17	13	Source	0.0012	.0255 (-)
TVOID	18	not entered	Source	not entered	.0004 (-)

S.I. FOR ALL OTHER PARAMETERS =0.000

* See foldout at end of report for detailed parameter listing

Table 4.5

PARAMETER IMPORTANCE RELATIVE TO DOSE VIA RADON INHALATION PATHWAY AT T=0 YEARS FOR SCENARIO 1

PARAMETER*	RANK		MODEL	INCREASE IN R^2	S.I. (INFLUENCE)
	S.I.	REGRESSION			
EMANAT	1	2	Radon Source Term	0.3947	.9778 (+)
MPCH2O	2	1	Radon Source Term	0.4511	.9077 (-)
F	3	4	Radon Inhalation	0.0365	.7143 (+)
CONCRA	4	3	Source	0.0431	.6667 (+)
OF	5	5	Radon Inhalation	0.0429	.5556 (+)
VDISP	6	7	Air Disp.	0.031	.3709 (-)
RHOTAL	7	6	Source	0.009	.3621 (+)
RAINGP	8	16	Source	0.0001	.1068 (-)
FWIND2	9	8	Air Disp.	0.0011	.0553 (+)
UWIND	10	9	Air Disp	0.0007	.0370 (-)
FACTCO	11	17	Source	0.0001	.0314 (+)
SULPRD	12	14	Source	0.0001	.0067 (+)
O2DIFF	13	not includ.	Source	-	.0042 (+)
NPRCP	14	13	Source	0.0001	.0018 (-)
KPRIME	15	11	Source	0.0003	.0012 (-)
HSRC	16	15	Air Disp.	0.0001	.0011 (-)
THFRAC	17	12	Source	0.0002	.0008 (+)
DISTCO	18	18	Source	0.0000	.0005 (-)
RAINTH	19	10	Source	0.0003	.0002 (-)

S.I. FOR REMAINING PARAMETERS EQUALS 0.000

* see foldout at end of report for detailed parameter listing.

Again the S.I. and stepwise regression techniques yielded similar results. The reader should note the dominance of the radium-related source term model parameters in the water consumption pathway.

The same method was applied to an importance analysis of pH at T=0 years as detailed in Table 4.6. By far the two most important parameters are sulphate production rate (SULPRD) and oxygen diffusivity (O2DIFF).

4.2.5 Scenario 1 Summary

Figures 4.10 and 4.11 indicate the contribution of each pathway to total dose at T=0 years and T=200 years, as well as the important parameters identified using the sensitivity index. The "important inputs" flowing into the "pathway boxes" are the five most important parameters for each individual pathway (based on the sensitivity index). The parameters listed between the "pathways boxes" and the "total dose box " are the 7 or 8 most important parameters involved in the calculation of total dose. The parameters noted are those which were important in both the calculation of total dose and also as inputs to the individual pathways.

4.3 Scenario 2: Pyrite Reduction

Scenario 2 assumes that a pyrite content of 0.6% in the uppermost metre of tailings compared to the 6.5% pyrite content for Scenario 1. All other characteristics of the tailings site, environmental pathways and receptor are the same as for Scenario 1.

Table 4.6

PARAMETER IMPORTANCE RELATIVE TO pH AT T=0 YEARS FOR SCENARIO 1

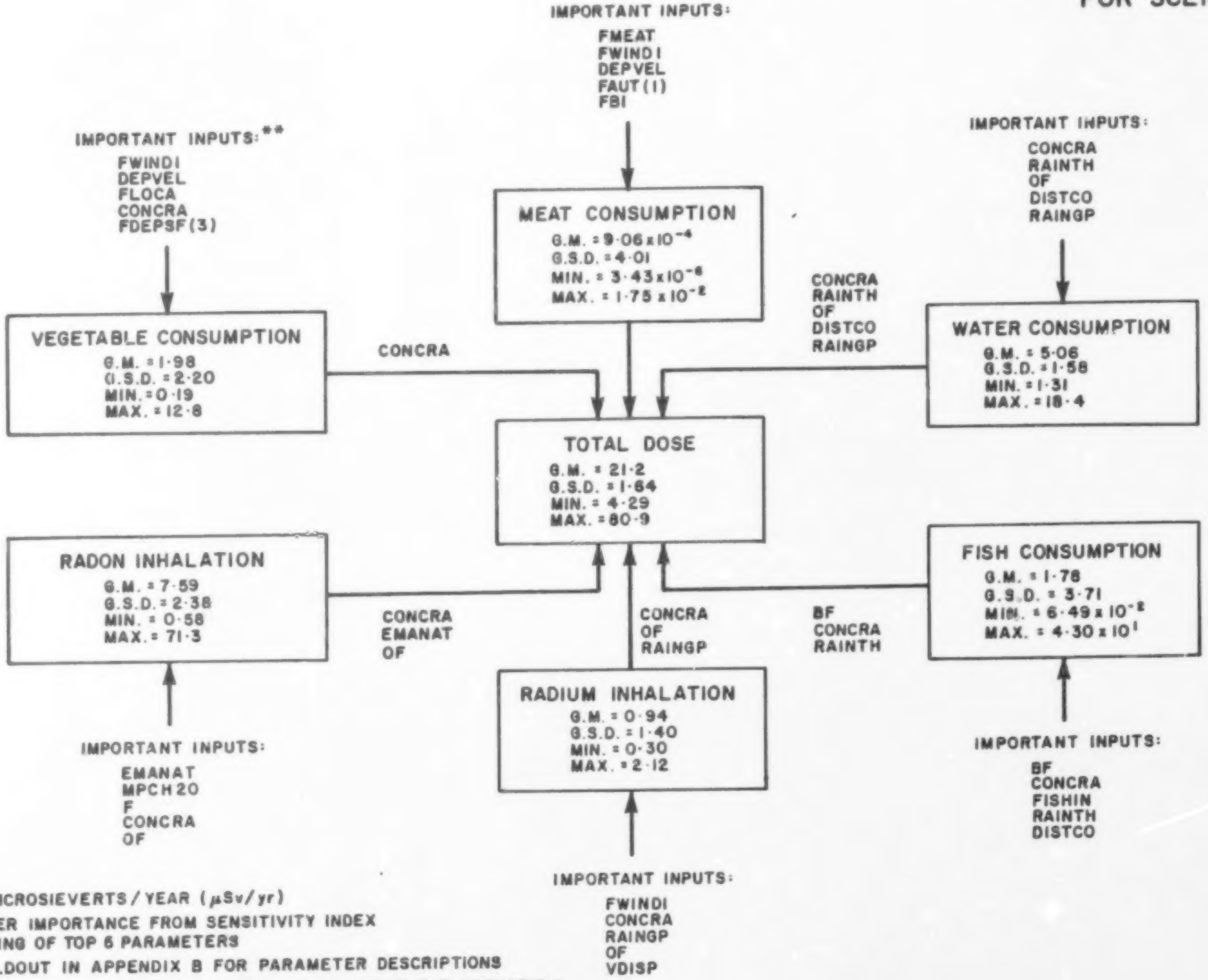
	RANK				
PARAMETER*	S.I.	REGRESSION	MODEL	INCREASE IN R^2	S.I. (INFLUENCE)
SULPRD	1	1	Source	.5940	.0269 (-)
O2DIFF	2	2	Source	.3200	.0169 (-)
KPRIME	3	4	Source	.0085	.0052 (+)
RHOTAL	4	3	Source	.0202	.0051 (-)
FACTCO	5	5	Source	.0041	.0018 (-)
NPRCP	6	6	Source	.0023	.0015 (+)

S.I. FOR ALL OTHER PARAMETERS =0.0

* See foldout at end of report for detailed parameter listing

FIGURE 4·10

CONTRIBUTION OF PATHWAYS TO TOTAL DOSE*
AND IDENTIFICATION OF IMPORTANT PARAMETERS WHEN T = 0 YEARS***
FOR SCENARIO I

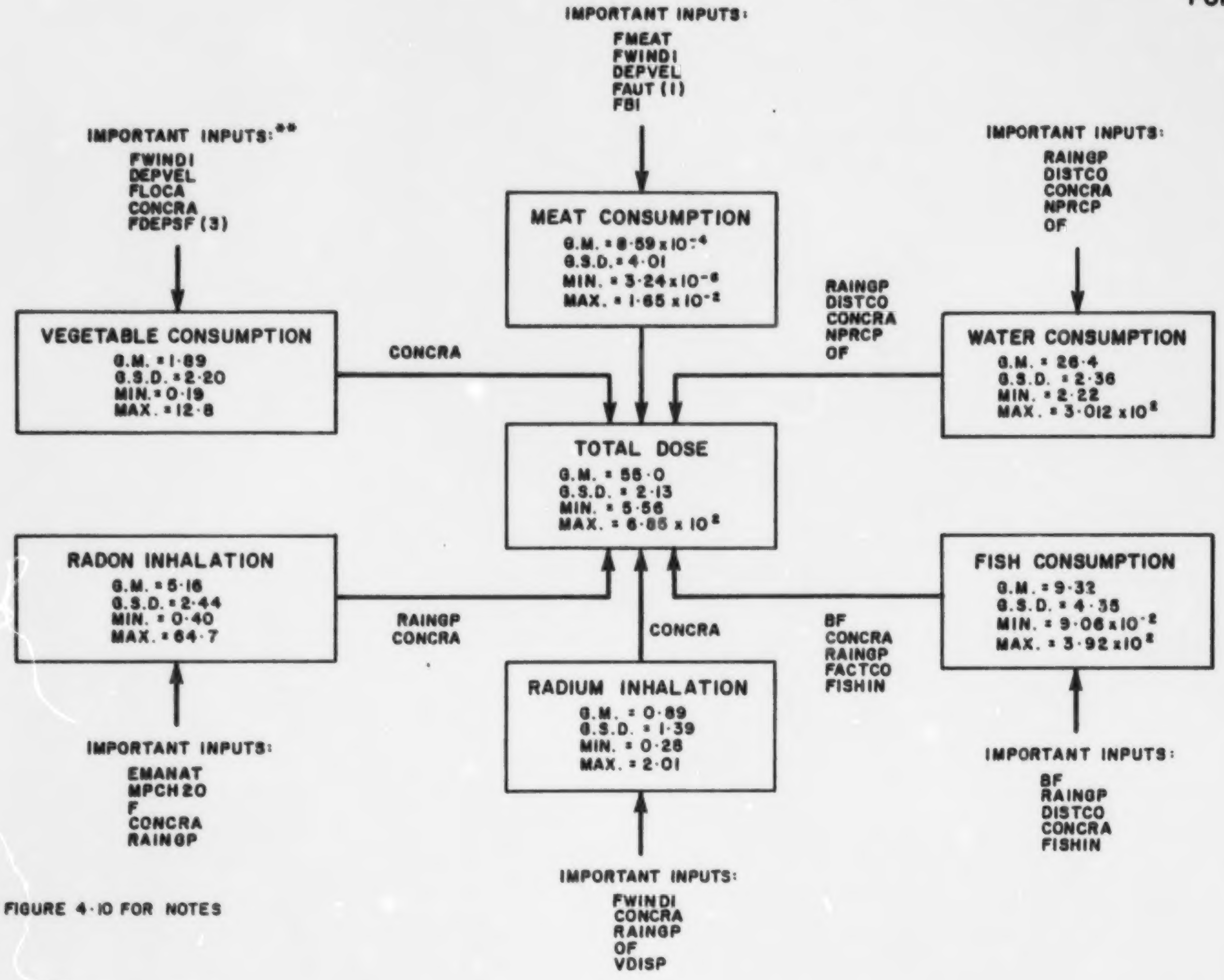


NOTE:

1. (*) DOSE IN MICROSIEVERTS/YEAR ($\mu Sv/yr$)
2. (**) PARAMETER IMPORTANCE FROM SENSITIVITY INDEX SHOWING RANKING OF TOP 6 PARAMETERS
3. (***) SEE FOLDOUT IN APPENDIX B FOR PARAMETER DESCRIPTIONS
4. ALL NUMERIC INFORMATION FROM 500 TRIAL SIMULATION FOR SCENARIO I

FIGURE 4·11

CONTRIBUTION OF PATHWAYS TO TOTAL DOSE*
AND IDENTIFICATION OF IMPORTANT PARAMETERS WHEN T = 200 YEARS***
FOR SCENARIO I



NOTE:

1. SEE FIGURE 4·10 FOR NOTES

4.3.1 Overall Evaluation

The time behaviour of the system with respect to total dose and the six contributing exposure pathways is shown in Figure 4.12. Comparison to the corresponding figure for Scenario 1 (Figure 4.2) indicates that the total dose from Scenario 2 rises to a higher peak at an earlier time (approximately 100 years) than Scenario 1 (which peaks at 200 years). Further comparison reveals that the air pathways for both scenarios remain approximately the same. The difference between the two scenarios thus occurs in the water pathways, which peak higher and earlier in Scenario 2.

This behaviour is physically reasonable. As discussed previously gypsum ($CaSO_4$) and radium sulphate ($RaSO_4$) form a solid solution in equilibrium with tailings porewater. High sulphate levels will drive the equilibrium toward the solid phase. In Scenario 1, with a 6.5% pyrite content, sulphate ions were in plentiful supply from the oxidation of pyrite. In Scenario 2, a much lower pyrite content (0.6% pyrite) is assumed. Consequently, the concentration of dissolved sulphate is much lower, leading to relatively rapid dissolution of the solid solution of gypsum and radium. Thus, in Scenario 2, a relatively rapid increase in dose (due to a rapid increase in the radium level in the lake) is observed followed by a correspondingly rapid decline in the same calculated parameters.

As could be anticipated, the doses for all pathways at T=0 years were the same for the two scenarios. Therefore, the dose histogram for total dose at T=0 years for Scenario 1 adequately describes Scenario 2. However the different peak at T=100 years for Scenario 2 is of interest and is presented in histogram format in Figure 4.13. Note the extreme skewness of the results and the substantially different geometric mean dose for Scenario

FIGURE 4·12

DOSE vs. TIME FOR SCENARIO 2
GEOMETRIC MEAN OF 500 TRIALS

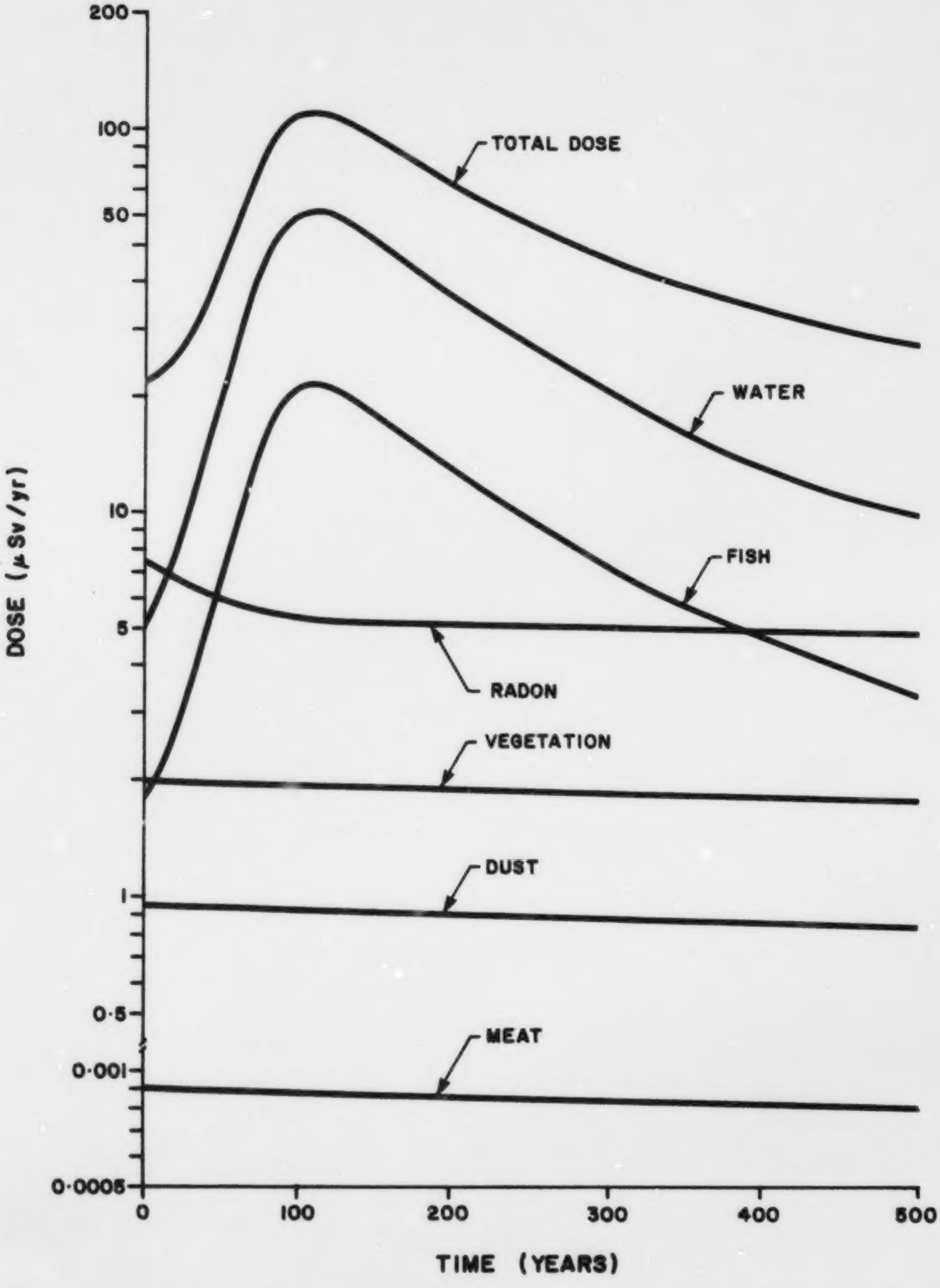
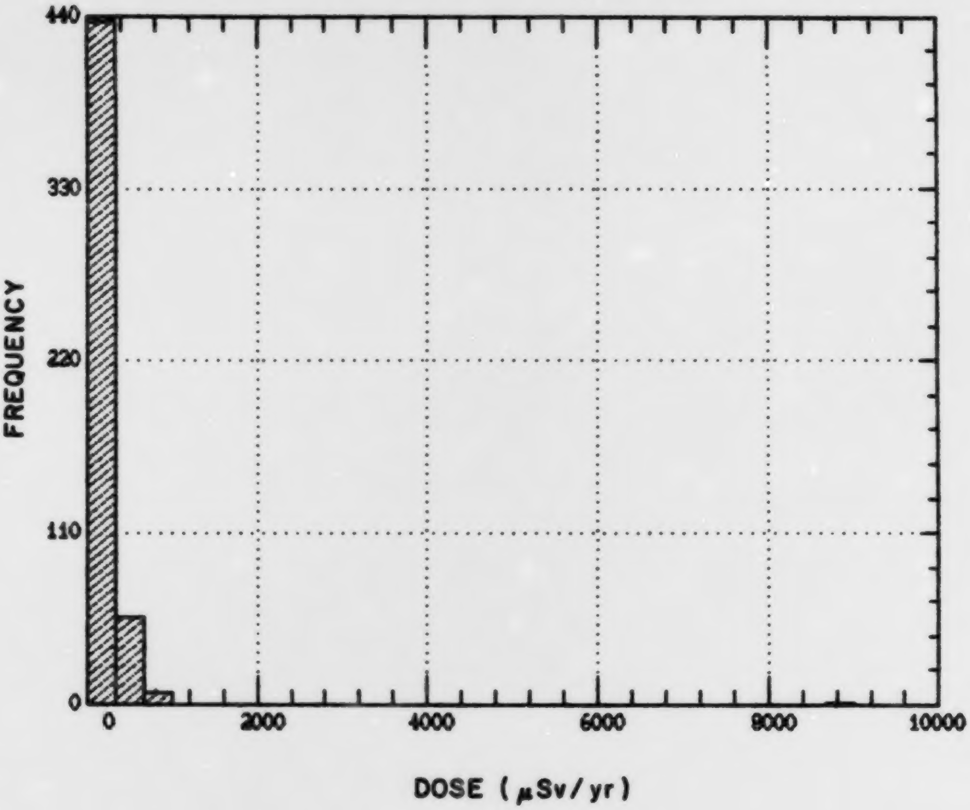


FIGURE 4·13

HISTOGRAM OF TOTAL DOSE AT T = 100 FOR SCENARIO 2



SUMMARY STATISTICS:

MEAN = 180
STANDARD DEVIATION = 414
GEOMETRIC MEAN = 111
GEOMETRIC STANDARD DEVIATION = 2·51
MINIMUM = 13·1
MAXIMUM = 8770

2 at T=100 years relative to that for Scenario 1 at T=200 years (shown in Figure 4.4b).

Importance analysis was performed on the results of Scenario 2, to investigate the effects of changing scenarios on the parameter importance rankings. Importance analysis of total dose at T=0 years for Scenario 2 yielded essentially the same results as for Scenario 1; indeed there was no distinguishable difference between the 2 regression analyses. Hence the importance analysis for total dose at T=0 years for Scenario 1 (represented in Table 4.1) applies directly to Scenario 2.

The results of importance analysis at T=100 years for Scenario 2 is shown in Table 4.7. Note the dominance of source term and water-related parameters, and the somewhat substantial difference between the results produced by regression analysis and sensitivity index analysis (for parameters ranked below the top 5). This illustates the desirability of applying several techniques for sensitivity and uncertainty analysis to a problem and drawing conclusions from the whole information base. Note also that the importance rankings for Scenario 2 (Table 4.7) show some similarities but differ in a number of respects from those for Scenario 1 (Table 4.2). Although the analyses were done for different points in time, both correspond to the time of maximum total dose, and thus should provide a valid basis for comparison. The results indicate that the importance analysis for a given scenario cannot be generalized to other scenarios. Each set of models and input parameters requires a unique importance analysis to provide a valid ranking of important parameters.

Full printouts of the stepwise regression analysis and the sensitivity index for total dose are provided in Appendix A.

Table 4.7

PARAMETER IMPORTANCE RELATIVE TO DOSE AT T=100 FOR SCENARIO 2

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	RANK					
PARAMETER*	S.I.	REGRESSION	DESCRIPTION	MODEL	INCREASE IN R^2	S.I. (INFLUENCE)
KSUBDR	1	1	Ra liquid-solid transfer coefficient (mixing zone)	Source	0.6483	.9691 (-)
BF	2	2	water to aquatic produce transfer coefficient: fish flesh	fish uptake	.0734	.9278 (+)
RAINGP	3	4	fraction of total Ra-226 which is associated with gypsum	Source	.0605	.8896 (+)
CONCRA	4	3	initial solid phase activity of Ra-226 in tailings	Source	.0603	.6665 (+)
SULPRD	5	37	sulphate prediction rate	Source	.0006	.5968 (+)
NPRCP	6	45	net precipitation	Source Lake	.0002	.5732 (+)
FACTCO	7	11	activity coefficient for divalent ions	Source	.0022	.4585 (-)
OF	8	5	Fraction of year in location	Inhalation (particulates & radon) Water Consumption	.0328	.4487 (+)
O2DIFF	9	19	oxygen diffusivity through tailings	Source	.0013	.4475 (+)
RHOTAL	10	21	tailings bulk density	Source Radium Source Term	.0015	.3138 (+)
MPCH20	19	6	percent moisture content weight of tailings	Radon Source Term	.0096	.0108 (-)
TVOID	31	7	porosity of tailings	Source & Radon Source Terms	.0054	.0007 (-)
VEGLAM(2)	-	8	pasture grass weathering decay constant	Vegetation Deposition	.0032	0.0
VEGLAM(1)	-	9	lichen weathering decay constant	Vegetation Deposition	.0041	0.0
HRSC	-	10	height of the source	Air Disp.	.0033	0.0

The time behaviour of pH is shown in Figure 4.14. Comparing it to the same graph for Scenario 1 (Figure 4.3) indicates that while both start at the same low value (4.0), the pH in Scenario 2 reaches equilibrium (pH=7.3) much more readily than Scenario 1 (50 years as opposed to 250 years). This is directly related to the much lower pyrite content of the tailings which limits its acid-generating ability. The much reduced variance of pH in Scenario 2 is also related to the reduced pyrite content.

Importance analysis of pH for Scenario 2 using sensitivity index and stepwise regression gave the same parameter rankings as in Scenario 1 (see Table 4.6 for Scenario 1 importance analysis of pH).

4.3.2 Scenario 2 Summary

Figure 4.15 indicates the contribution of each pathway to total dose at T=100 years as well as identifying important parameters via the sensitivity index. The methods used to identify parameters important to specific pathways and to the overall calculations were outlined in the previous section. The major difference between Scenarios 1 and 2 is that Scenario 2 peaks earlier (at 100 years) and higher (Dose = 111 uSv/yr) than Scenario 1 (which peaks at approximately 180 years with a dose of 55 uSv/yr). Different behaviour of the water-related pathways accounts for the difference in time behaviour of the two scenarios. The reduced pyrite level in Scenario 2 accounts for the differences in both total dose and pH.

4.4 Scenario 3: Pyrite Reduction and Limestone Addition

This scenario assumes that the pyrite content of the tailings has again been reduced from 6.5% to 0.6%. In addition, limestone is assumed to have been incorporated in the depyritized tailings.

FIGURE 4-14

pH vs TIME FOR SCENARIO 2

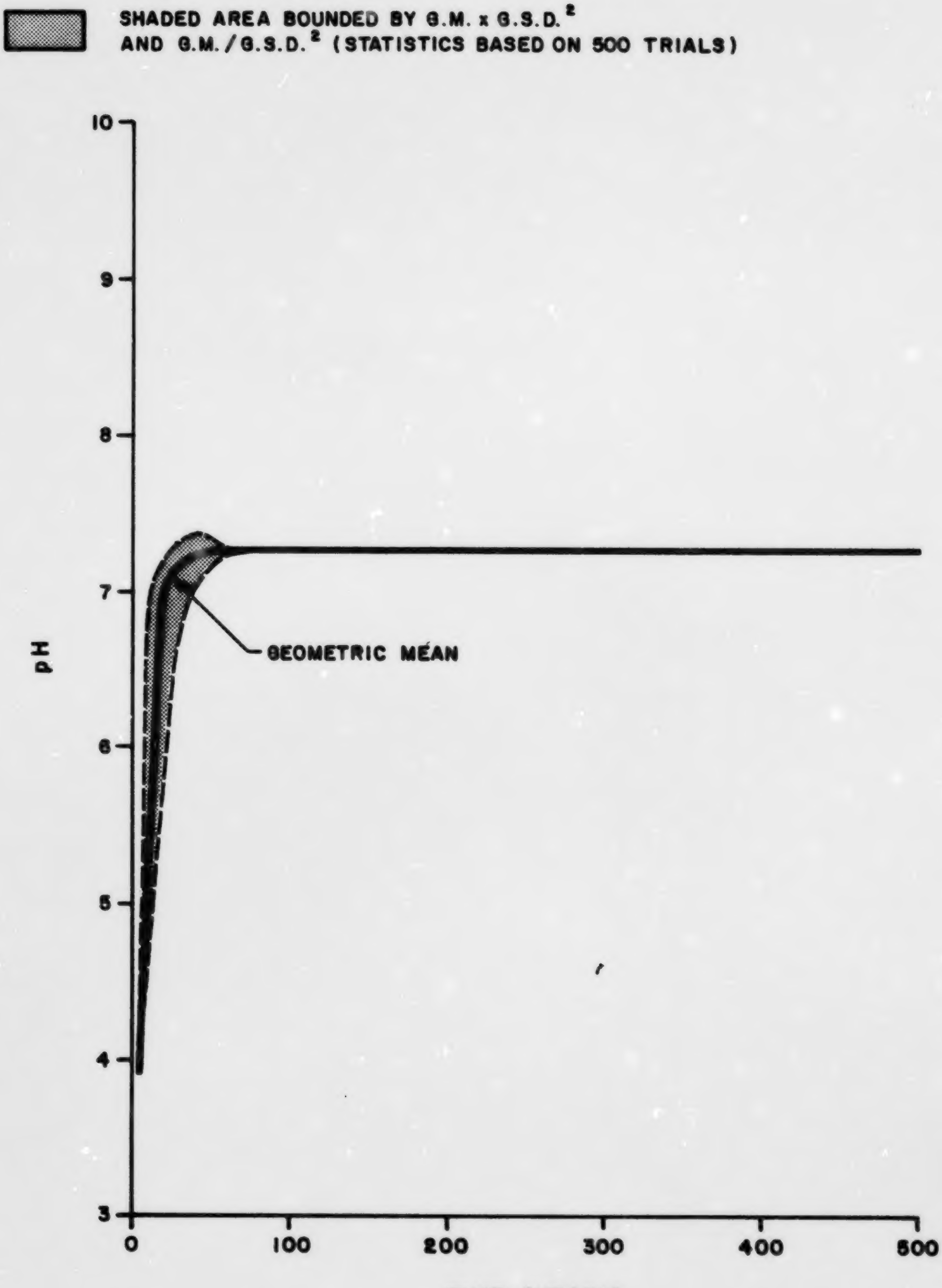
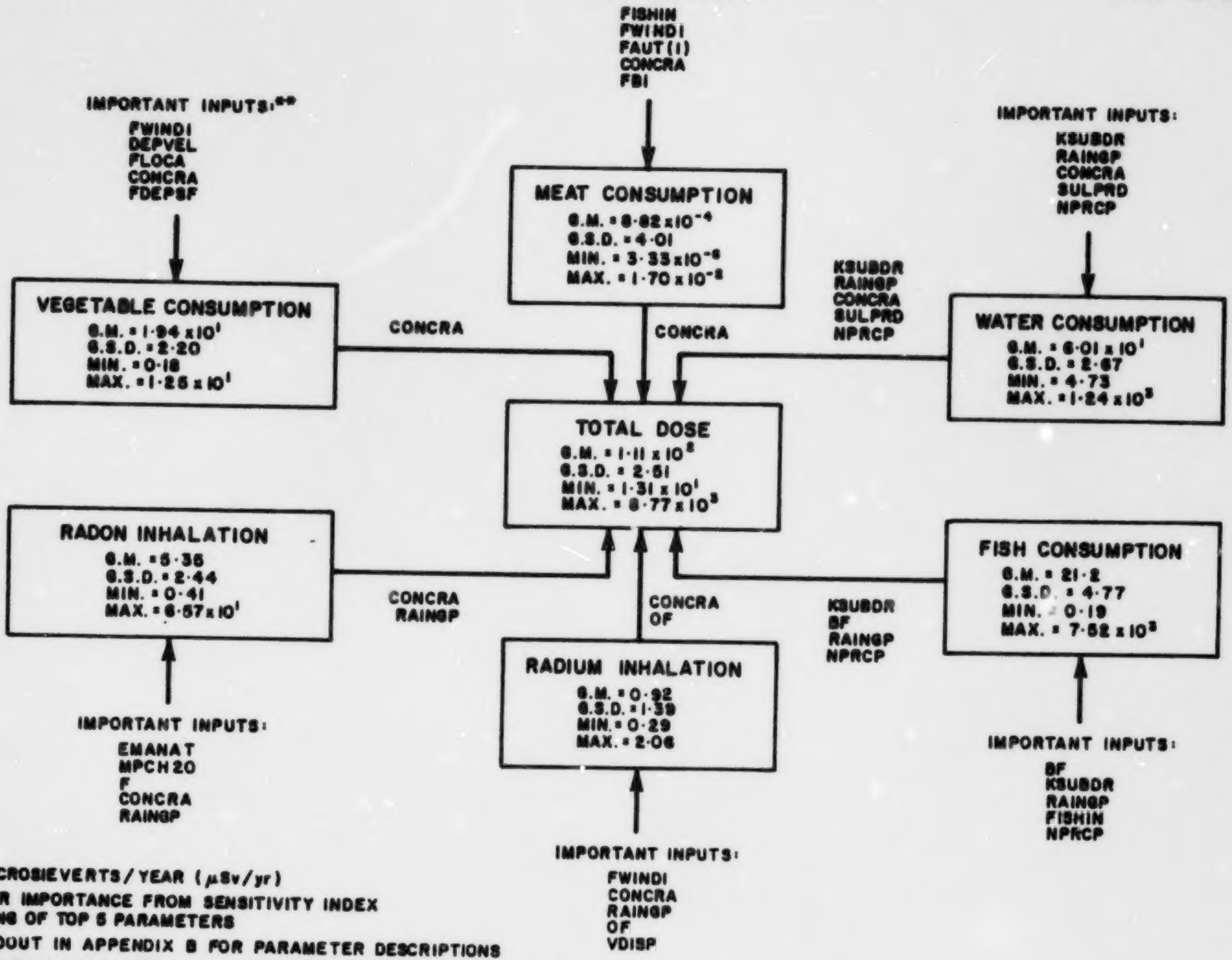


FIGURE 4·15

CONTRIBUTION OF PATHWAYS TO TOTAL DOSE*
AND IDENTIFICATION OF IMPORTANT PARAMETERS WHEN T = 100 YEARS***
FOR SCENARIO 2



NOTE:

1. (*) DOSE IN MICROSIEVERTS/YEAR (μSv/yr)
2. (**) PARAMETER IMPORTANCE FROM SENSITIVITY INDEX SHOWING RANKING OF TOP 5 PARAMETERS
3. (***) SEE FOLDOUT IN APPENDIX B FOR PARAMETER DESCRIPTIONS
4. ALL NUMERIC INFORMATION FROM 500 TRIAL SIMULATION FOR SCENARIO 2

The 1% (by weight) $CaCO_3$ (lime) added to the tailings represents an excess over that required for stoichiometric equivalence with pyrite. Limestone addition alters several of the chemical equilibria in the tailings, hence a number of changes were required to the source term models. These are described in Appendix A of Volume 3. None of the other component models were changed.

4.4.1 Overall Evaluation

The time-varying behaviour of the receptor dose arising from the six pathways is presented in Figure 4.16. This can be compared to the corresponding figures for Scenarios 1 and 2 (Figures 4.2 and 4.12). As with Scenario 2, the dose arising from radon and the airborne radium pathways is unaffected. The results indicate that the dose via all water pathways (and hence the total dose) is initially higher than in the previous two scenarios. The total dose continues to remain somewhat above that for Scenario 2 throughout the 500 year modelling period.

The principal factors which influence the behaviour of the total dose curve are the same as those noted in Scenario 2. The addition of limestone, which maintains the tailings porewater pH near neutral, further retards the pyrite oxidation rate, resulting in lower concentrations of dissolved sulphate ions. The dissolution of the calcium/radium sulphate coprecipitate is thus enhanced, leading to higher dissolved radium levels. The higher dissolved radium concentration in the tailings porewater is reflected in a higher dose via the water pathways.

Histograms of the total dose at T=0 and at T=100 years (the time of maximum mean dose) are presented in Figures 4.17 (a) and (b). These can be compared with histograms of total dose arising from the other scenarios (Figures 4.4 and 4.12).

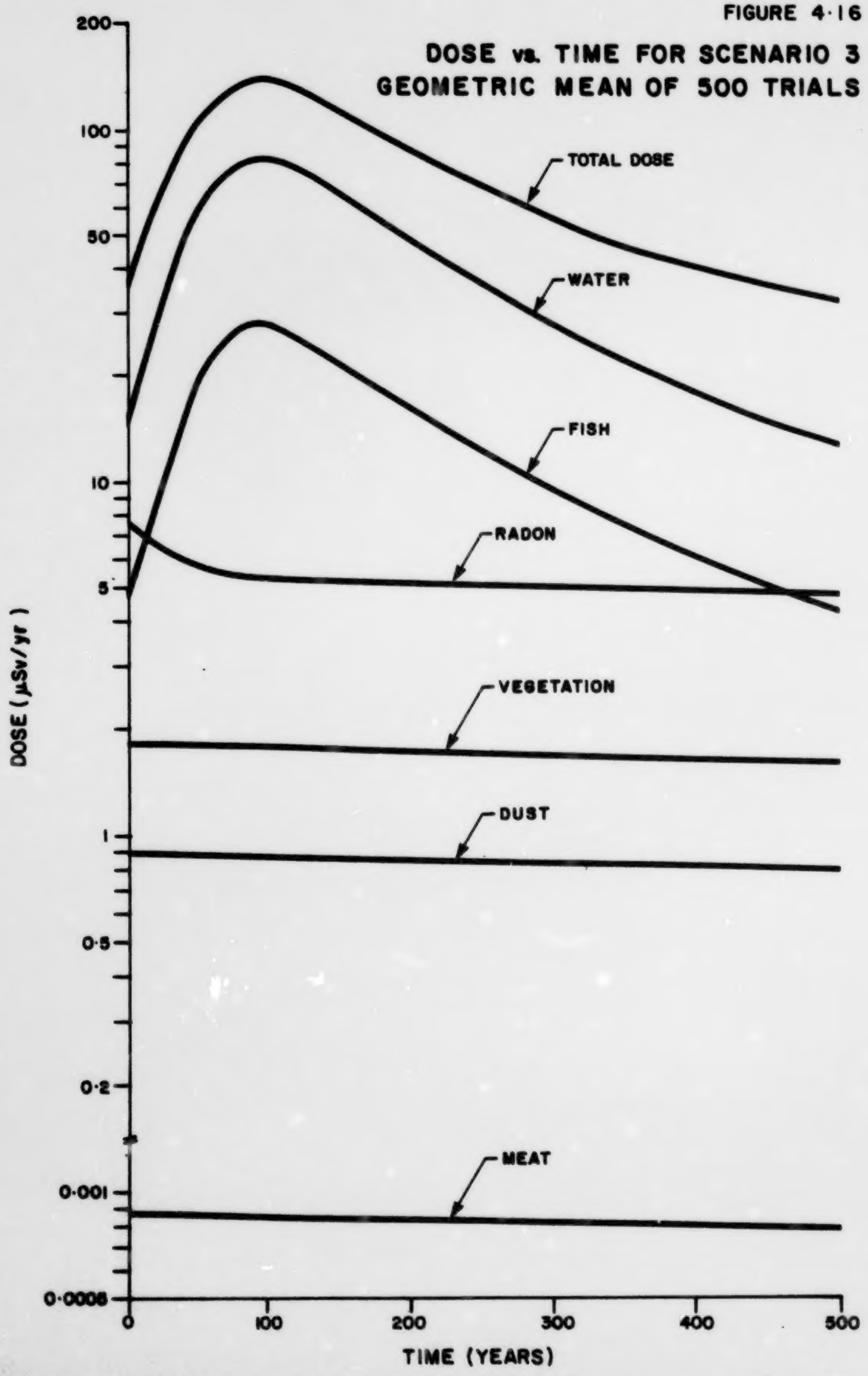
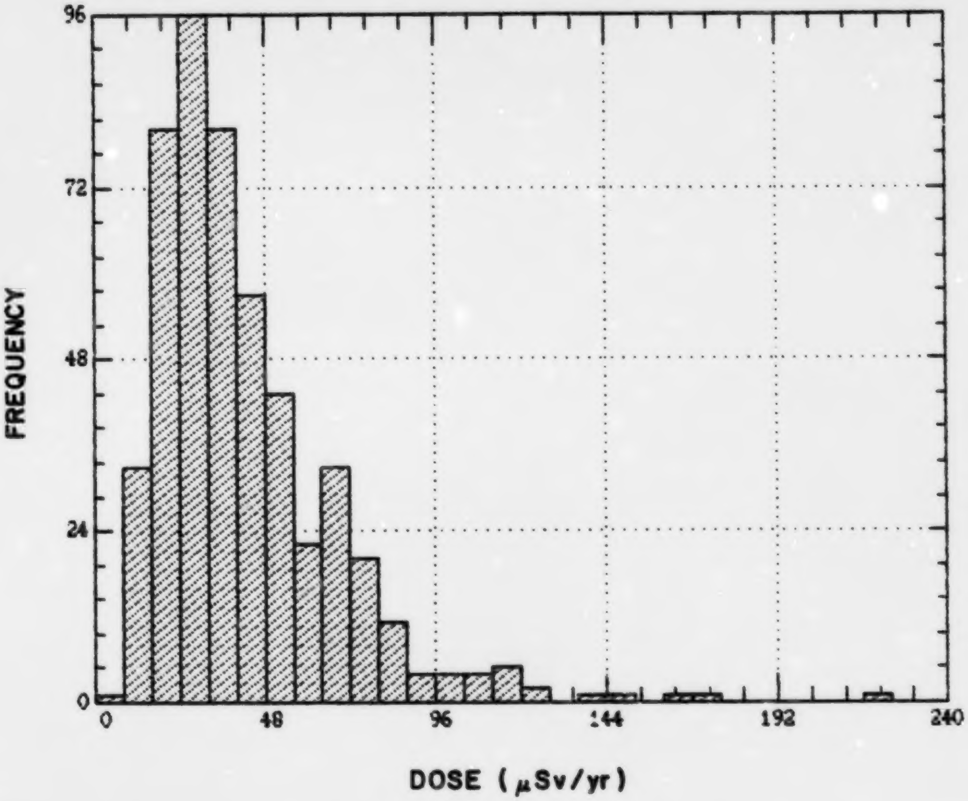


FIGURE 4·16

DOSE vs. TIME FOR SCENARIO 3
GEOMETRIC MEAN OF 500 TRIALS

FIGURE 4·17a

HISTOGRAM OF TOTAL DOSE
AT T=0 FOR SCENARIO 3

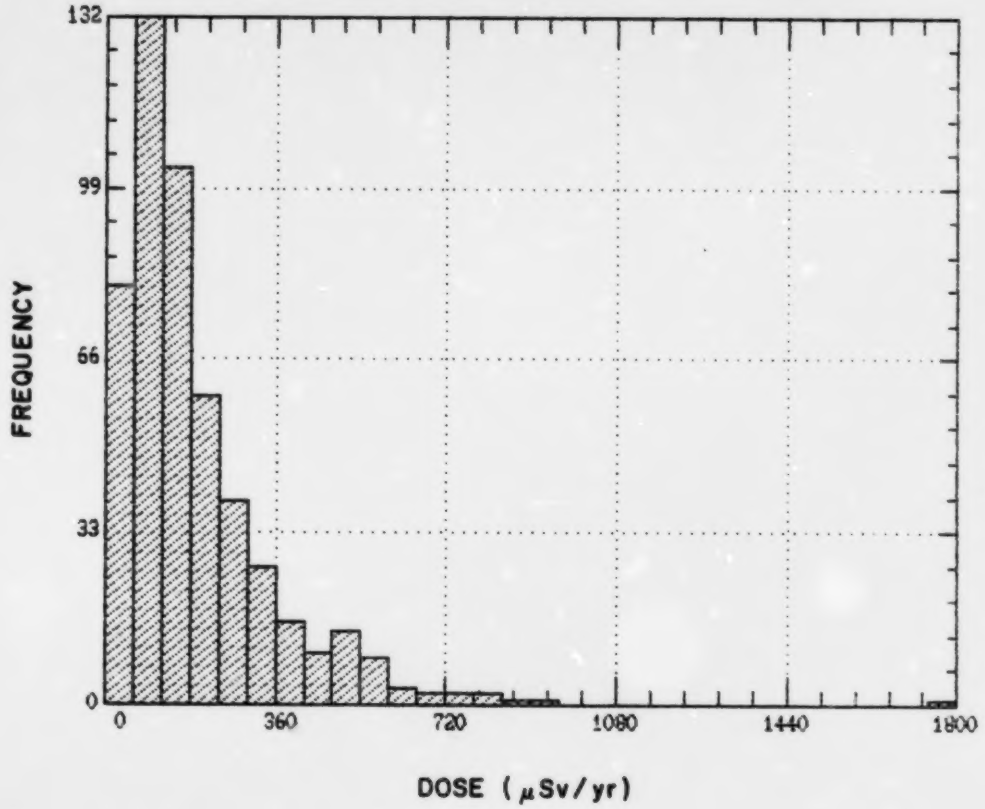


SUMMARY STATISTICS:

MEAN = 42·5
STANDARD DEVIATION = 25·8
GEOMETRIC MEAN = 36·4
GEOMETRIC STANDARD DEVIATION = 1·74
MINIMUM = 7·60
MAXIMUM = 221

FIGURE 4·17b

HISTOGRAM OF TOTAL DOSE AT T = 100 FOR SCENARIO 3



SUMMARY STATISTICS:

MEAN = 187·5
STANDARD DEVIATION = 168·4
GEOMETRIC MEAN = 136·8
GEOMETRIC STANDARD DEVIATION = 2·2
MINIMUM = 15·8
MAXIMUM = 1755

The time-varying behaviour of lake pH is presented in Figure 4.18. As expected, the addition of limestone as a buffering agent has produced a relatively consistent, near-neutral pH. The increased concentration of Ca^{2+} and CO_3^{2-} contribute in large part to the higher pH in the lake. A small decrease in mean pH is noted between 100 and 150 years. This is due to the depletion of the added limestone and an attendant shift in chemical equilibria. The increased width of the "error bands" in this interval reflects several complexities, including the fact that limestone depletion occurs at varying times and pyrite oxidation occurs at varying rates for each of the 500 trials. The behaviour of the system between years 200 and 500 shows no variability in the mean pH. This observation is due to the fact that the natural background concentrations of the principal ionic species were input as constant values in the pH model.

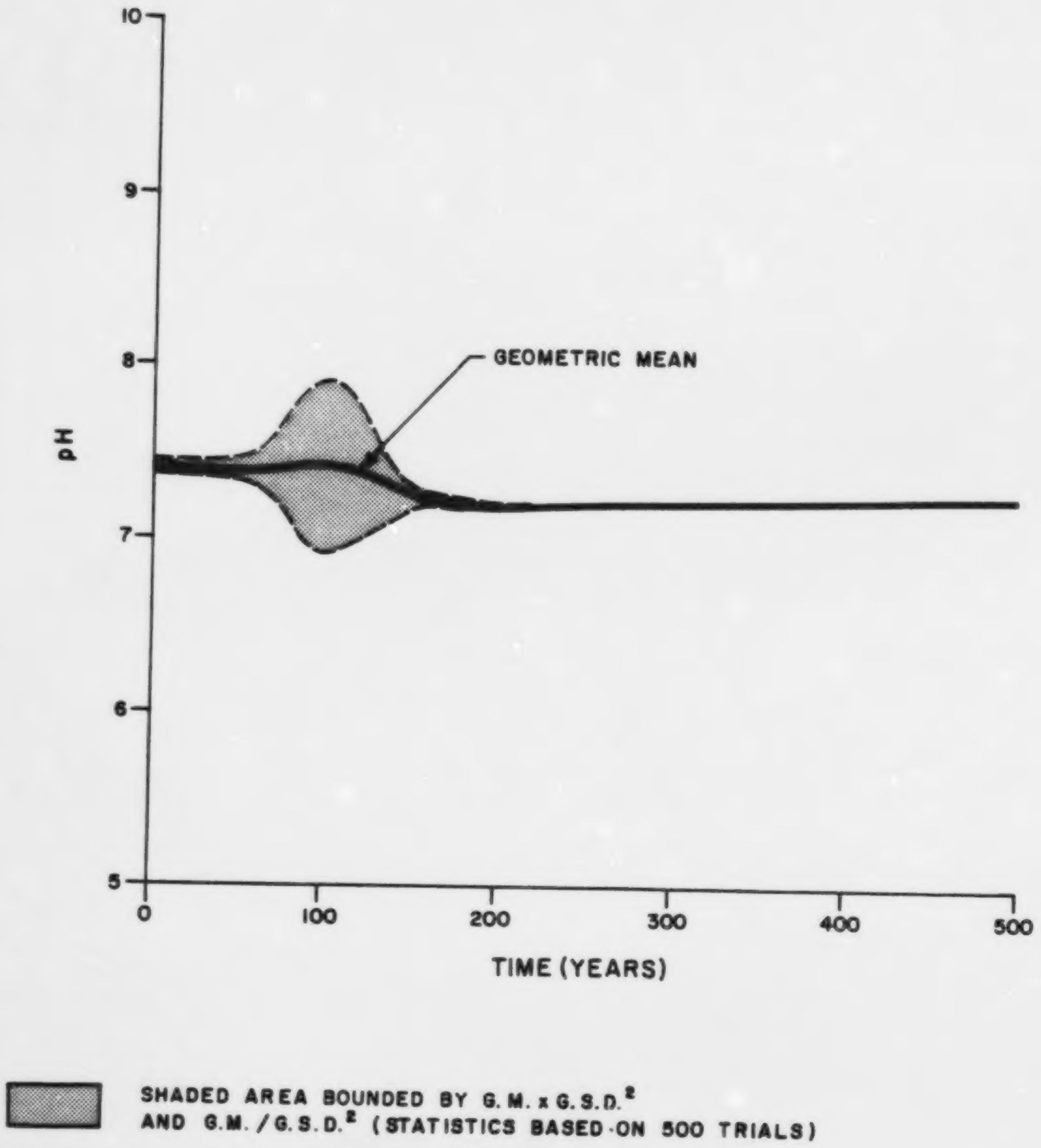
In summary, the combination of pyrite reduction and limestone addition produces slightly greater radiological consequences but a much more stable (and desirable) pH response, than the previous two scenarios. An importance analysis was not performed for this scenario.

4.5 Scenario 4: Bulk Float and Limestone Addition

As described previously, bulk flotation involves separation of a significant fraction of the radium and thorium from tailings during the milling process. It is assumed that 25% (by weight) of the tailings are removed via this process, the gypsum ($CaSO_4$) content is reduced from 5% to 0.5% and the radium and thorium levels are reduced by about 70%. Again, limestone addition is assumed in this scenario. Although the relative concentrations of the various species are altered in this scenario, the physical and chemical processes are assumed to be identical to those

FIGURE 4-18

pH vs. TIME FOR SCENARIO 3



▨ SHADED AREA BOUNDED BY G.M. x G.S.D.^{2}
AND G.M./G.S.D.^{2} (STATISTICS BASED ON 500 TRIALS)

modelled in Scenario 3.

4.5.1 Overall Evaluation

Once again, the time-varying behaviour of dose to the reference receptor is presented graphically (see Figure 4.19). As expected from the reduction in radionuclide (radium & thorium) levels, the dose over much of 500 years lies below that for the "base case" (see Figure 4.2 for comparison). The slightly higher initial total dose is due to higher initial doses via the water pathways as explained below.

The limestone addition and reduced pyrite content enhances the dissolution of the calcium/radium coprecipitate. This leads to a higher initial dissolved radium concentration, even though the radionuclide inventory is reduced by bulk flotation. As expected, this initially produces higher doses via the water pathways. Note that as a result of radionuclide removal, the receptor dose via radon and all airborne pathways is substantially lower than in the previous three scenarios. Histograms for total dose at T=0 years and T=50 years (the time of maximum mean dose) are presented in Figures 4.20 (a) and (b).

The lake pH again shows very little change over time. As indicated on Figure 4.21, the (mean) pH is initially about 7.4 and stabilizes after 50 years at 7.2. This behaviour is expected because of the effects of the limestone addition described earlier. The error bands follow the trend noted in the previous scenario although the time frame is shortened because of the reduction in gypsum levels. An importance analysis was not performed for this scenario.

FIGURE 4·19

DOSE vs. TIME FOR SCENARIO 4
GEOMETRIC MEAN OF 500 TRIALS

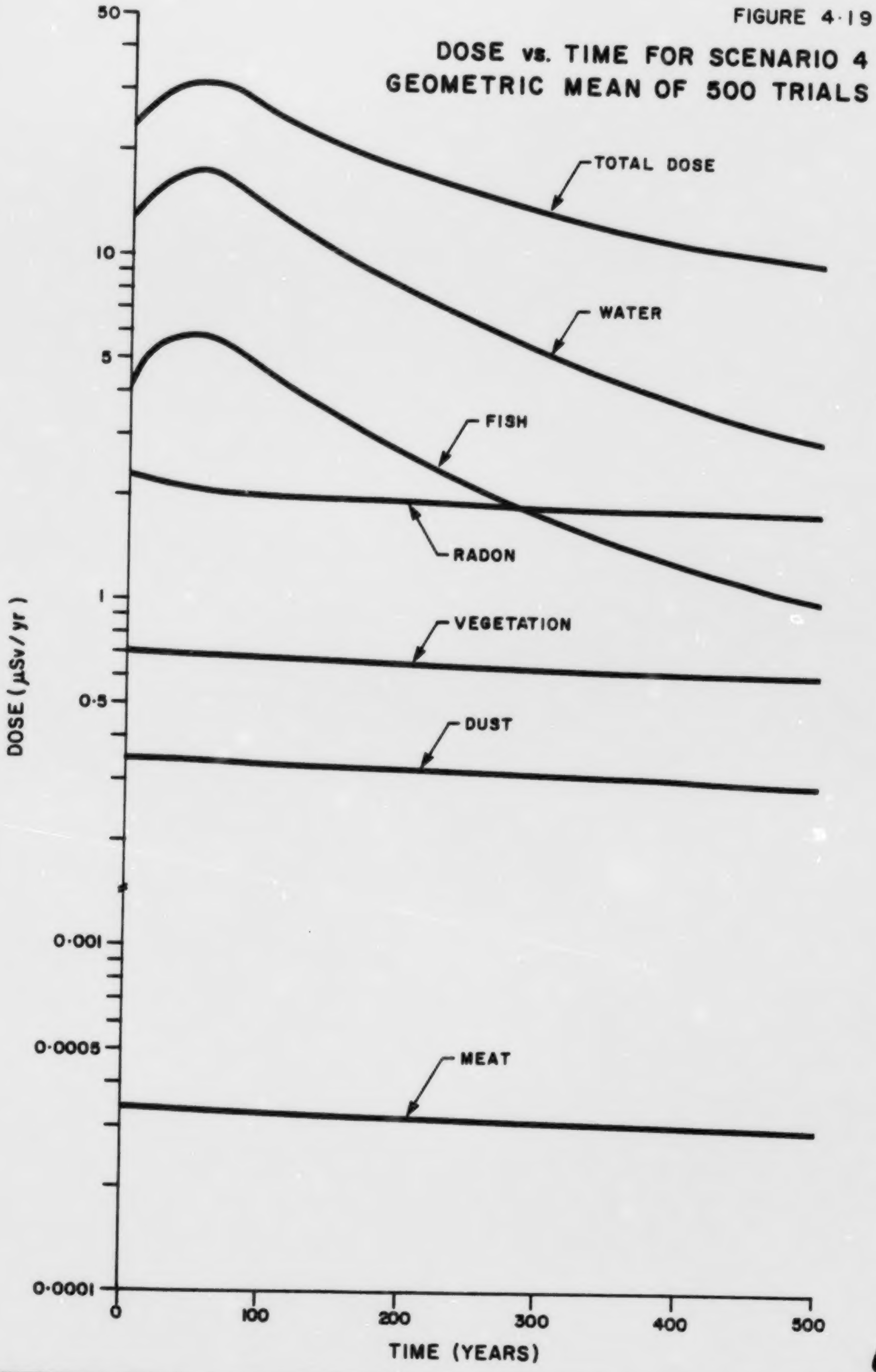
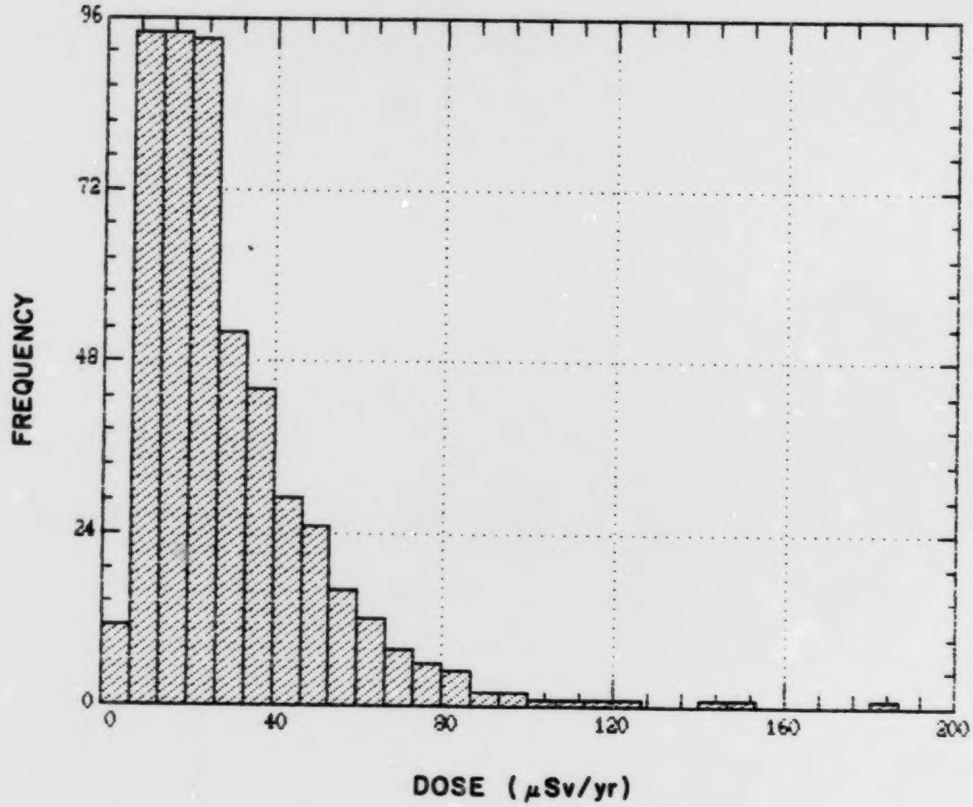


FIGURE 4·20a

HISTOGRAM OF TOTAL DOSE AT T=0 FOR SCENARIO 4

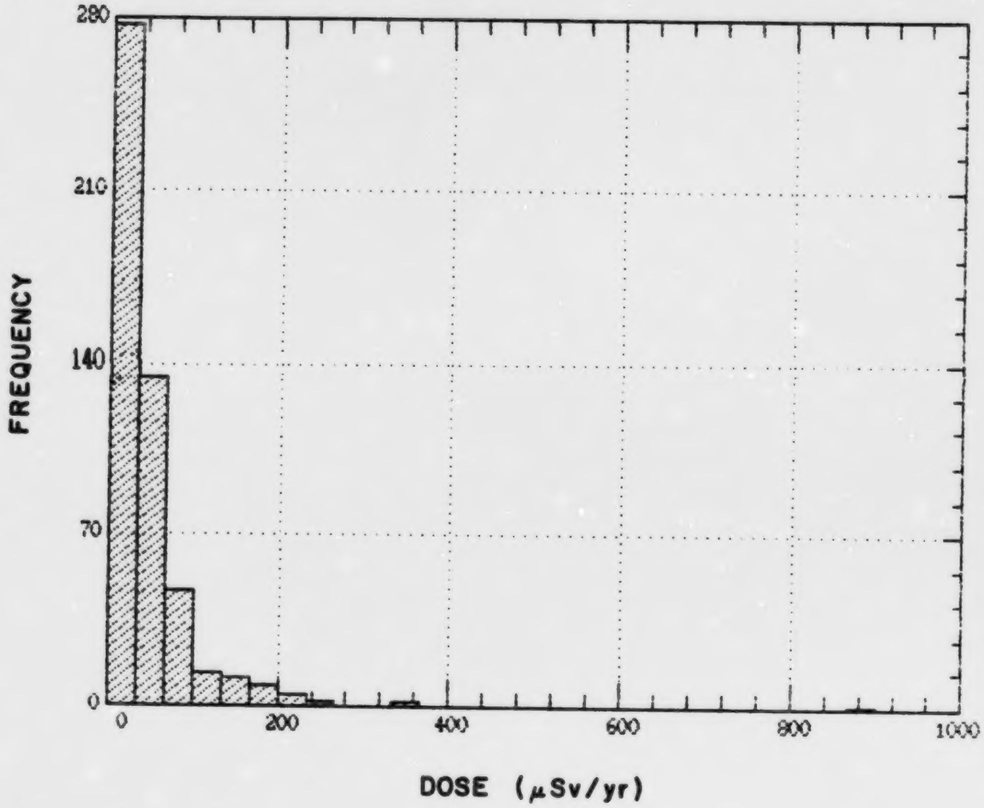


SUMMARY STATISTICS:

MEAN = 29·7
STANDARD DEVIATION = 22·1
GEOMETRIC MEAN = 23·7
GEOMETRIC STANDARD DEVIATION = 1·95
MINIMUM = 3·48
MAXIMUM = 185

FIGURE 4·20b

HISTOGRAM OF TOTAL DOSE
AT T=50 FOR SCENARIO 4

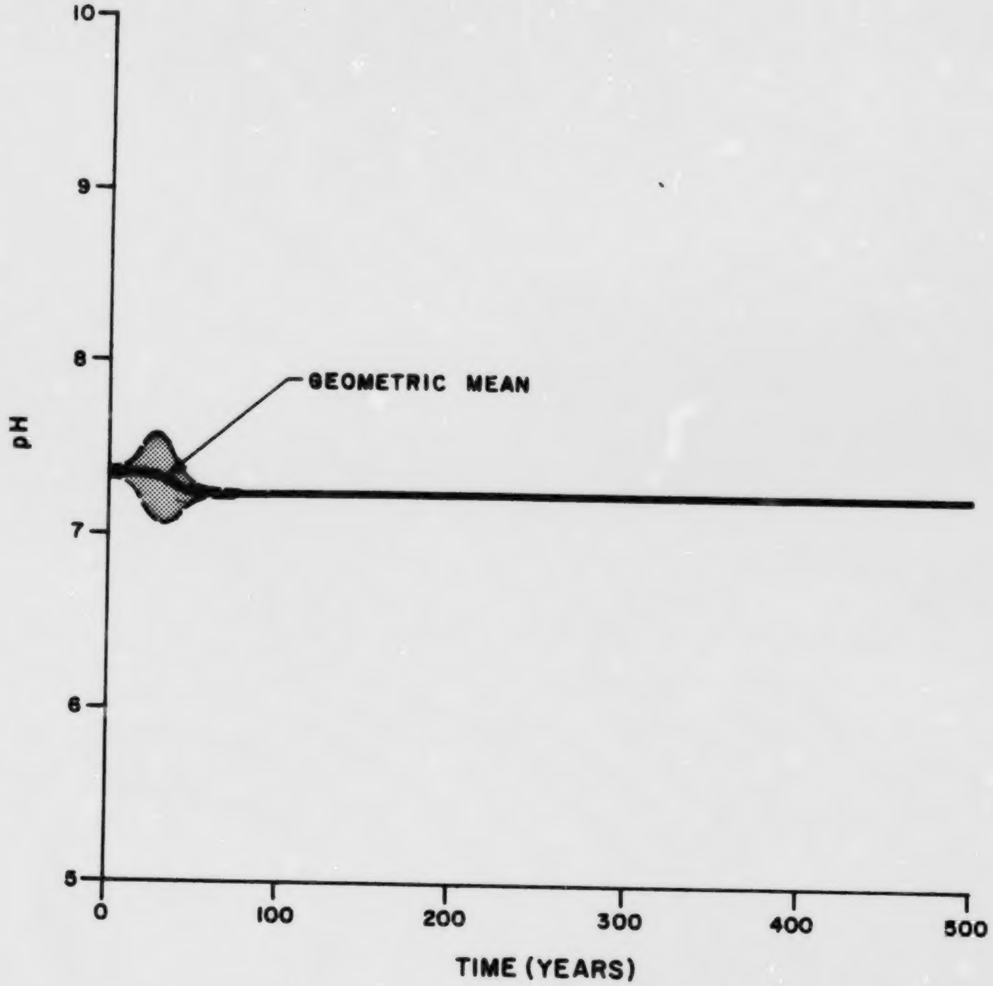


SUMMARY STATISTICS:

MEAN = 29·7
STANDARD DEVIATION = 22·1
GEOMETRIC MEAN = 23·7
GEOMETRIC STANDARD DEVIATION = 1·95
MINIMUM = 3·48
MAXIMUM = 185

FIGURE 4·21

pH vs. TIME FOR SCENARIO 4



▨ SHADED AREA BOUNDED BY G.M. x G.S.D.^{2} AND G.M. / G.S.D.^{2} (STATISTICS BASED ON 500 TRIALS)

4.6 Summary of Alternative Management Options

A considerable amount of information has been presented in the previous sections. Much of the salient information is summarized graphically in Figures 4.22 and 4.23. The time-varying behaviour of total receptor dose for each of the 4 scenarios is presented in Figure 4.22. Figure 4.23 presents similar curves for lake pH.

Figure 4.22 indicates that pyrite reduction (Scenarios 2 & 3) which is paralleled by an increase in the mass of gypsum and radium dissolved results in higher total doses. Pyrite removal is implicit in the bulk flotation process (Scenario 4), which also results in decreased inventories of gypsum, radium and thorium. The reduced radium inventory dissolves more quickly, producing a higher short-term dose than in the base case. The peak dose for Scenario 4 however is below the base case peak. The dose rapidly falls to below the base case level and remains there for the remainder of the simulation period. Interpretation of the dose information presented on Figure 4.22 must be tempered by the realization that the dose curves for Scenarios 2,3 and 4 do not include the effects of disposal of the pyrite concentrates.

As indicated on Figure 4.23, lake pH is greatly affected by both pyrite reduction (Scenario 2) and by limestone addition (Scenarios 3 and 4). Pyrite reduction does not affect the initial pH, but causes it to rise to a steady state value much more quickly than in the base case. Limestone addition, as expected, results in a relatively stable, near-neutral pH throughout the simulation period.

FIGURE 4·22

TOTAL DOSE vs. TIME SUMMARY
GEOMETRIC MEAN VALUES OF 500 TRIALS

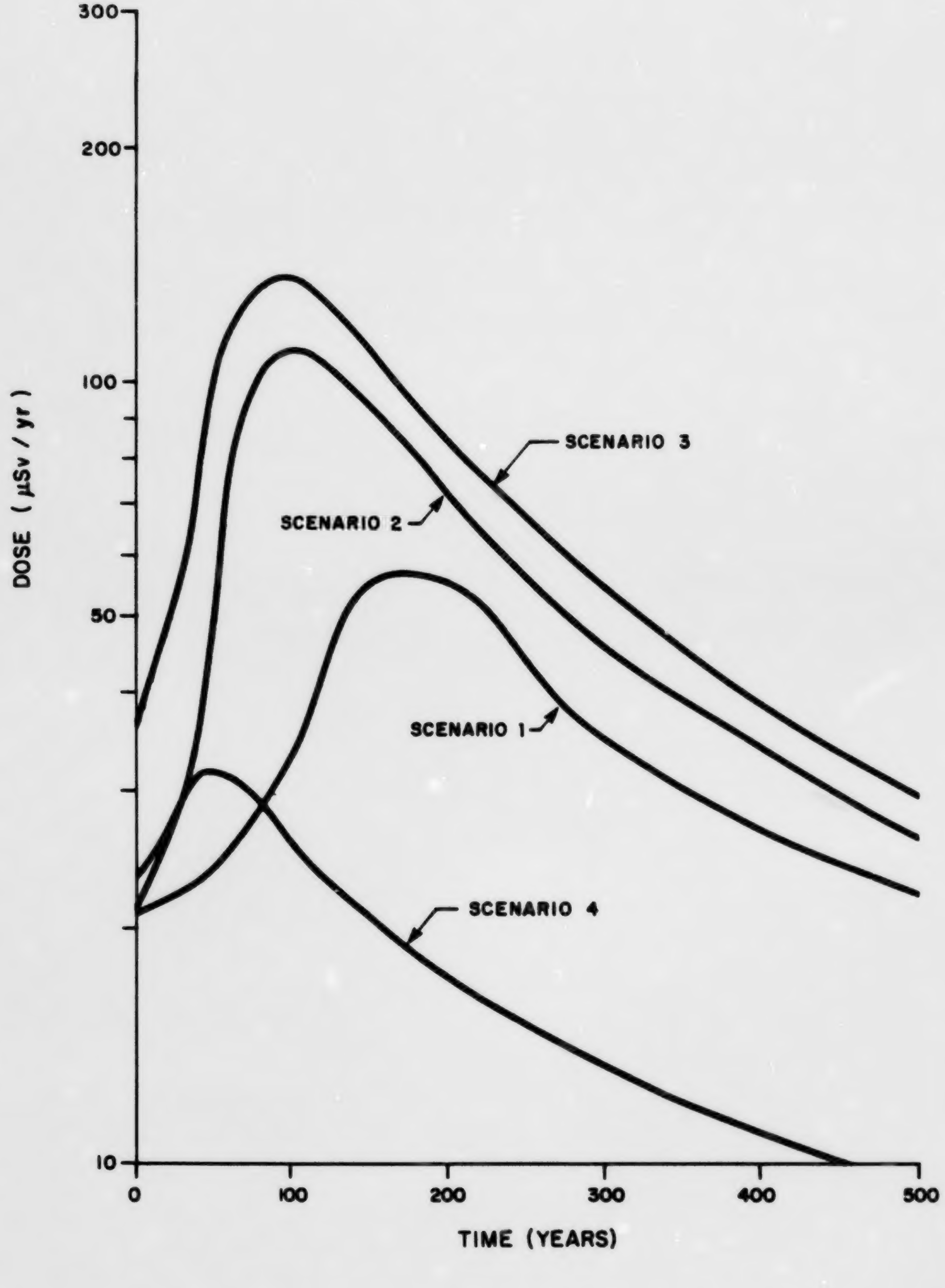
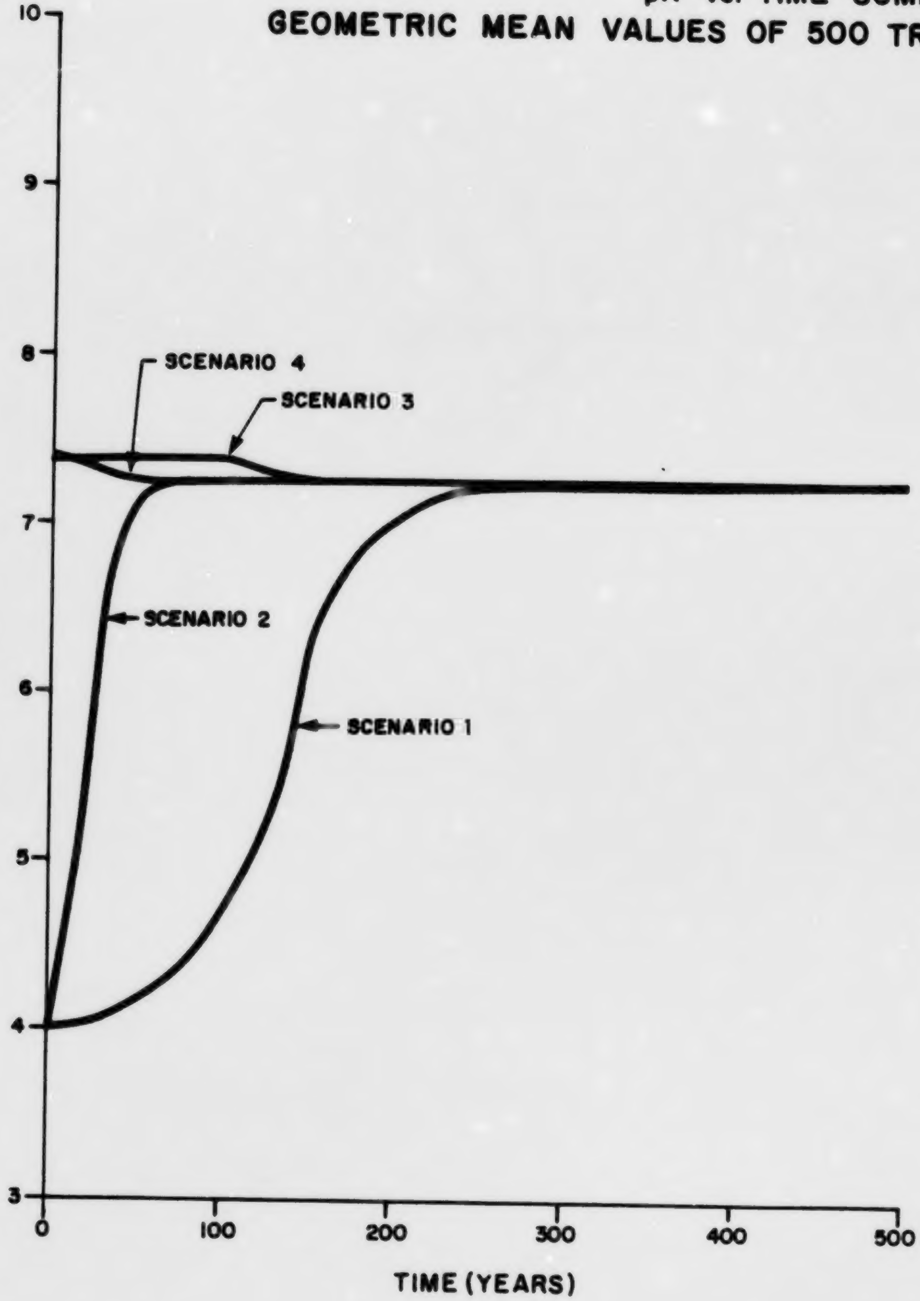


FIGURE 4-23

pH vs. TIME SUMMARY
GEOMETRIC MEAN VALUES OF 500 TRIALS



5.0 SUMMARY OBSERVATIONS FROM TOPICAL DISCUSSION PAPERS

The preparation of topical discussion papers proved most helpful in generating a technical exchange among the study participants. Some of the topical papers prepared during the Phase II studies were required in order to permit the Phase II modelling to proceed. Others were innovative attempts to investigate technical areas potentially of great significance to carrying out or evaluating probabilistic assessments.

The following discussion papers were prepared as part of the Phase II studies and are presented as the appendices to Volume 2 of this report.

Appendix A	The Interpretation of a Probabilistic Assessment of the Environmental Effects of Uranium Mill Tailings by S. Kaplan, L.M. Lowe, and D.B. Chambers
Appendix B	Probabilities of Individual Cells or Tail Area Under Monte Carlo Sampling by P.M. Reilly
Appendix C	Confidence Regions for Output Distributions from Monte Carlo Calculations by P.M. Reilly and T.A. Duever
Appendix D	Sensitivity/Uncertainty Analysis and Evaluation of Time Response of a Model of Radionuclide Accumulation in Soil by D.I. Beals, D.B. Chambers and B.E. Halbert
Appendix E	Techniques for Handling Correlated Parameters by T. Andres and B. Goodwin

Appendix F	Modelling Radium Exchange Between the Water Column and Lake Sediment by J.M. Scharer and B.E. Halbert
Appendix G	A Dynamic Model of Radium-226 Behaviour in an Oligotrophic Lake by R.H. Gardner, B.G. Blaylock and F.O. Hoffman
Appendix H	The Data Base Available for Estimating Radium-226 Bioaccumulation in Fish Flesh by B.E. Halbert, D. Stone and V.J. Cassaday
Appendix I	The Data Base Available to Develop Parameter Distributions for use in Modelling the Transfer of Airborne Radioactivity to Terrestrial Vegetation by V.J. Cassaday, B.E. Halbert and D. Stone
Appendix J	The Transfer of Radium from the Atmosphere to Terrestrial Vegetation by F.O. Hoffman, R.H. Gardner and B.G. Blaylock

The papers by Halbert et al (Appendix H) and Cassaday et al (Appendix I) provided the basis for certain parameter values used in the Phase II modelling exercise.

The paper by Scharer and Halbert (Appendix F) describes a model of radium exchange between lake water and sediment. This model was required for the investigations into the dynamics of an oligotrophic lake (Appendix G) and for use in the Phase II environmental simulation (lake dispersion module).

Appendix E prepared by Goodwin and Andres provides a convenient summary of techniques to handle parameter correlation in environmental simulations.

Two topical papers (Appendices G and J), address various aspects of model aggregation in the aquatic and terrestrial environment respectively.

The remaining topical papers (Appendices A,B,C,D) all discuss several technical issues of importance to carrying out and interpreting the results of probabilistic analysis.

There is little merit in repeating by rote the details of the technical papers one after another. Instead it was considered most appropriate to discuss a small number of selected technical issues on a topic by topic basis (as opposed to a paper by paper basis). These summary discussions form the remainder of this chapter.

One issue of considerable technical importance, which is addressed in the discussion paper by Beals et al (Appendix D) and to a lesser extent in the technical paper by Kaplan et al (Appendix A) is the issue of stepping through time. This aspect is not discussed in this chapter since it was addressed earlier in the report (Section 2.6).

The reader interested in further details regarding any topics discussed here is encouraged to examine the pertinent topical paper(s) in Volume 2.

5.1 Some Aspects of Parameter Correlation

The credibility of a probabilistic analysis depends, in part, on how well the models and data used in the simulations are accepted by reviewers. One requirement for plausibility and acceptance is that values of parameters input to the simulation model display the same characteristics and relationships as do values observed in the field, laboratory and literature. In particular, observations that are known to be correlated in the field should correspond to parameter values that are correlated in the model.

In the most general sense, correlation exists when there is a mutual relationship between two variables. The correlation coefficient is a measure of the (linear) relationship between two variables. The correlation coefficient always lies between +1 and -1, either value indicating there is an exact linear relationship between the variables. Positive values of the correlation coefficient indicate a tendency for the two variables to increase together. Negative values indicate that large values of one variable are associated with small values of the other variable. A correlation coefficient of zero means there is no relationship between the variables.

There are several reasons for incorporating parameter correlations into mathematical models. First, some real world observations are known to be correlated, and hence should, for the sake of realism, be modelled as such. Secondly, by specifying correlations in a model, one is less likely to generate unrealistic scenarios involving unlikely or implausible combinations of input variables when employing sampling techniques. Finally, by using correlations in sampling parameters, one may gain further insight into how various parameters directly or indirectly affect simulation results.

Given that including known correlations between model parameters is required for a long-term assessment, methods are needed for ascertaining correlations. The problems in developing such methods are partly scientific (how to measure correlations in the field) and partly psychological (how to question experts to obtain consistent and unbiased information about correlations expected). The scientific questions can be tackled by carefully-designed research programs, involving the simultaneous measurement of a number of variables. Where correlations are a matter of theory or expert opinion, rather than experimental data, care must be taken in eliciting this information. A number of techniques have been proposed for properly eliciting expert opinion.

Several techniques exist to generate correlated parameter values for use in probabilistic analysis. Three of these are described briefly below and are discussed in more detail in the paper by Andres and Goodwin (Appendix E of Volume 2). As will be noted, the techniques deal primarily with pairwise correlation. One method is designed to (approximately) incorporate a multiple correlation matrix.

Where the form of a relationship between parameters is known or suspected, a functional relationship may be the easiest way to express the dependence. A functional relationship is a formula that expresses one parameter as a function of other, more basic parameters. One of the benefits of developing functional relationships among parameters is the opportunity provided to put components of the model to experimental test. It is much easier to generate evidence to reject a faulty relationship than it is to reject a correlation coefficient. A model that has been stringently tested and survives will be more credible than one that is based on theoretical considerations alone.

Should one wish to specify sets of variables as truly correlated rather than functionally related, there are methods available by which it may be accomplished. As Reilly (1985) pointed out it is important to recognize that there is an infinite number of ways in which two or more variables may be combined to give a required correlation structure and at the same time preserve required marginal distributions. One well-known procedure described by Reilly to assure uniqueness of the result involves specifying that all variables in the set be normally distributed. It is then possible to specify any feasible correlation structure amongst any number of random variables, and to sample them accordingly by a simple method.

The technique of Kleijnen (1974) achieves uniqueness in generating a sample from a pair of correlated variables by preserving the entire distribution of one variable, the mean and variance of the second variable, and the correlation coefficient between the two. The algorithm is quite straightforward and is presented in the paper by Andres and Goodwin (in Appendix E of Volume 2). It has been incorporated into the system code as part of the RANSIM computer program, although it was not used to specify parameter correlations in the Phase II work.

Also, Iman and Conover (1981) have proposed a method for generating data sets which incorporate multivariate correlations. It involves the use of rank correlation and matrix manipulation and is considerably more involved than Kleijnen's technique. It does have the advantage of preserving the distribution of all parameters, and is not adversely affected by truncation (as Kleijnen's technique is). One drawback is that manipulation of the matrices involved can be time- and space-consuming. In the case of large simulations or limited computer resources these requirements may be prohibitive.

Reilly (1985) has described another method for generating a pair of correlated random variables which have a given coefficient of correlation and which preserves any given marginal distributions for the individual variables. It can easily be extended to any number of random variables. Uniqueness is achieved by specifying arbitrarily the sampling probabilities of a required number of joint probabilities of pairs of particular values of the variables.

5.2 Model Aggregation

Aggregation is defined in Funk and Wagnalls Dictionary as: "to bring or gather together, as in a mass, sum or body". In terms of modelling it refers to the mathematical lumping of complex multi-variate processes into simpler expressions.

The papers by Hoffman, Gardner and Blaylock in Volume 2 (Appendices G and J) present case studies of model aggregation for two physical processes, namely:

1) radium-226 behaviour in an oligotrophic lake, and
2) the transfer of radium-226 from the atmosphere to pasture vegetation

In the first paper, a multi-component model for predicting the dynamics of Ra-226 behaviour in an hypothetical oligotrophic lake situated in the Precambrian Shield region of Canada is described. Time-dependent solutions to the model are obtained for (1) a simulation of a clean lake with a contaminated input and (2) a simulation of a contaminated lake with a 99.9% cleanup of the input.

A simpler aggregate model was also developed, involving three components (water, sediment, game fish), and was found to predict adequately the dynamic behaviour of the lake. At equilibrium, Ra-226 concentrations predicted by the multi-component dynamic model are not appreciably different from those predicted using the empirically-derived, aggregate model.

In the second paper, three models, of varying complexity, for predicting the concentration of Ra-226 in pasture vegetation from continuous deposition of atmospheric particulates containing Ra-

226 are described. Two of the models include processes describing the interception of wet and dry deposition by the surfaces of vegetation and the root uptake of Ra-226 from the soil horizon. The third model accounts only for direct interception of particulate deposition. Uncertainties in the model parameters were included in the analysis through the use of probability distributions. The models were evaluated probabilistically to assess the effects on the predicted vegetation concentrations for two scenarios: (1) after 50 years of continuous deposition; and (2) after equilibrium is established between the atmosphere and soil. The results show that the most complex model has the largest uncertainty. However, all three models predict comparable mean concentrations in the pasture vegetation for the scenarios examined. While data are not available to assess which of the three models is most suitable, it can be argued that the third model (direct interception only) should be used for assessment purposes when there is a continuous source of airborne particulate radium-226.

It was further suggested that the larger uncertainty in the complex model may have been the result of a failure to account for parameter correlation in the model. Parameter correlations tend to be accounted for implicitly in aggregate models. Also more complex models tend to have more parameters with their associated distributions (uncertainties) which tend to broaden the output distribution.

In both examples, the complex and aggregated models produce similar mean value results. Given that a complex model and an aggregated model of the same phenomena yield similar answers, the fundamental issue in model aggregation is the tradeoff between the detailed information about the system provided by the complex model and the simplicity of the aggregated model. Complex models are desirable for studying and understanding system dynamics but a much simpler model can, in many cases, be used to make reliable predictions. For assessment purposes, where everyone from the scientist to the general public will be reviewing the models it is suggested that the simplicity offered by model aggregation is preferable to more complex models. Model aggregation is particularly desirable in the early stages of a study, in order to pinpoint critical pathways, providing the focus for the study and more complex modelling efforts in these pathways.

5.3 Sensitivity/Uncertainty Analysis Techniques

The paper by Beals et al in Appendix D of Volume 2 details the development of a simple model describing radionuclide movement in a soil column. The model was evaluated probabilistically and was subjected to an investigation utilizing a variety of sensitivity/uncertainty analysis techniques. The techniques ranged from the relatively complex techniques of regression analysis to conceptually simpler techniques including the sensitivity index, uncertainty index and a graphical technique. The investigation was designed to evaluate the applicability/appropriateness of the various methods for determining (ranking) the important parameters input to a simulation model.

Regression analysis uses the output of a number of trials in order to fit a linear response surface to the simulation model. The sensitivity index for a particular parameter is essentially the global derivative of that parameter, taken with all other parameters held constant at their nominal values. The uncertainty index directly relates model input variance to output variance. The graphical technique involves plotting the change in model response with respect to changes in one parameter with all other parameters held at their nominal values. Interestingly, in the specific case of the soil model, all of the techniques yielded similar results. This observation was not always found to hold with other component models and it was concluded that the consistency was specific to the case of the soil model.

Sensitivity/uncertainty analysis of different models requires a flexible creative approach, and indeed is somewhat of an art. It is not a rote, automatic means of identifying sensitive and uncertain variables, but rather is part of a feedback loop between the sensitivity/uncertainty analysis, the model and the real system being modelled. It should stimulate questions that must be rationally answered rather than answering questions in an uncritical fashion.

The examples indicate that no single method of sensitivity/uncertainty analysis is ever likely to be best always, and in most cases it is desirable to apply several such tools. In the Phase II studies two tools were used for sensitivity/uncertainty analysis, namely stepwise regression and the sensitivity index. The use of these techniques was outlined in Section 2.3 and is discussed in detail in Appendix D of Volume 2.

The use of time-variant input parameters, (i.e. parameters such as net annual precipitation, which vary from year to year) confounds sensitivity/uncertainty analysis. In part, the problem occurs since, if one chooses to sample time-variant parameters in time steps shorter than the simulation length, problems arise in applying regression analysis. The following discussion briefly addresses the problem.

Consider a model with three non-time-variant parameters sampled at the beginning of each trial and one time-variant parameter sampled at every time step. Let us assume for discussion purposes a simulation time period ranging from 0 to 2 years in yearly steps. Hence for a given trial there would be:

i) one value for each of the three non-time-variant parameters,
ii) three values for the time-variant parameter, one for each time period (0, 1 and 2 years), and
iii) three calculated output values, one for each time period.

At a given time, for a specified number of trials, one could regress the set of outputs for the trials against the set of three non-time-variant inputs and the single time-variant input for that time. The hypothesis in performing the regression is that there is a direct relationship between the output and the inputs. However, the system could possess "memory" of the previous time steps, the effect of which would depend on the values of time-variant parameters sampled in previous time steps. Hence the interpretation of a regression analysis containing a time-variant input is not clear.

At present in UTAP the problem has been avoided by sampling time-variant input parameters from a narrowed distribution of mean values for the simulation length (i.e. long-term mean value selected from a suitably adjusted distribution of mean values). In essence, some fine details of the transient behaviour of the system have been sacrificed to facilitate sensitivity/uncertainty analysis.

The sensitivity index also breaks down in the presence of time-variant parameters, since it is predicated on evaluating the model over the entire simulation period with all parameters held constant.

It is felt that further investigation of the problems outlined above would lead to the development of techniques which would permit the modeller to investigate transient behaviour yet retain the ability to carry out more meaningful sensitivity/uncertainty analysis on input parameters.

5.4 Statistical Confidence in Monte Carlo Calculations

Probabilistic analysis refers to the process of drawing samples from one or more input distributions using a Monte Carlo technique and substituting these parameter values in simulation models to produce values of some calculated variable(s).

Often, the most important output from such simulations is the probability density function, obtained in approximate form as a histogram of calculated values. The sample frequency curves (or histograms) are not identical to the true, underlying population probability curves. However, they are expected to resemble the population probability curves more closely as the sample size (i.e. the number of trials or repetitions) increases.

For example, if the number of Monte Carlo trials were extremely large (say 10^9) and the number of histogram cells moderate (say 20), it would be expected that the relative frequency curve would be a smooth curve, and would agree very closely with the true probability density function. It would also vary little from one Monte Carlo run to the next if the same number of trials were done. If however only 100 Monte Carlo trials were made with 20 cells, the relative frequency diagram would likely be "ragged", and vary a good deal from one run of 100 trials to the next. This "raggedness" or "shimmer" arises from the finite number of samples associated with the Monte Carlo process.

We can then ask the question, "Given a Monte Carlo run with a stated number of trials, which produced a stated number of counts in a certain interval on the range of a calculated variable, how strongly may we believe (i.e. what is the probability) that the true value of the variable is in that interval?" It is also possible to ask the question "Given a Monte Carlo run with a stated number of trials, which resulted in a given histogram for a calculated variable, how confident are we (i.e. what is the probability) that a particular value of the variable will be exceeded on repeated sampling?

Reilly (Appendix B of Volume 2) and Kaplan et al (Appendix A of Volume 2) adopt a Bayesian approach to answering these questions. Both papers discuss the theoretical basis for techniques to address the above questions (taking somewhat different approaches) and illustrate the techniques by numeric examples.

In the Phase I study a distribution-free technique based on an extension of the Kolmogorov-Smirnov goodness-of-fit test (as described for example in Kendall and Stuart (1961)), was used to develop confidence bands for the cumulative distribution function. The test is based on the maximum absolute deviation between the empirical (i.e. observed) cumulative distribution function, and the underlying true cumulative distribution. Unfortunately, while the method is attractive because of its simplicity, it does have the disadvantage of being too conservative, particularly in the tails of the distribution. Usually, ordinate values of less than zero for the left tail and greater than one for the right tail are obtained. This consideration prompted the further investigation of techniques for establishing confidence bands which led to the topical paper by Reilly and Duever, presented as Appendix C of Volume 2.

The method discussed by Reilly and Duever is based on the work of Cheng and Iles (1983), in which a probability density function of known mathematical form is fitted to the observed relative frequency values (by estimating the values of the parameters in the probability density function). At the same time, by conventional statistical methods, a joint confidence region for the parameter values is found. Then, the set of mathematical functions which might, at the confidence level being used, include the true probability density function, is obtained by substituting into the known mathematical form of the probability density function all values of the parameters, within their joint confidence region. The envelope of this set of functional relations is the confidence region, at the confidence level being used, for the probability density function of the variable.

Reilly and Duever discuss the theoretical basis of the technique and extend it to the cumulative distribution function. The authors provide numeric illustrations of the technique which indicate that the confidence bands generated in this manner are substantially narrower than those calculated using the Kolmogorov-Smirnov approach.

6.0 SUMMARY AND RECOMMENDATIONS

6.1 Summary

The Phase II studies were designed to build upon the results of the preliminary probabilistic assessment carried out in Phase I. In general terms the objectives of the Phase II investigations were to:

- incorporate time-dependent behaviour,
- include additional radionuclides,
- include non-radioactive constituents,
- define parameter distributions for transfer factors,
- evaluate conventional disposal technologies,
- carry out a preliminary assessment of a generic site,
- demonstrate the utility of sensitivity/uncertainty analysis techniques and,
- continue development of the system code.

In addition, a number of special technical topics were investigated in Phase II including:

- interpretation of probabilistic results,
- methods for estimating confidence regions on output distributions,
- techniques for handling correlated parameters,
- model response to time-variant input parameters and,
- implications of model aggregation.

The incorporation of time-dependent behaviour was a particularly important aspect of the Phase II investigation. To this end, it was necessary to develop time-dependent component models as well as to restructure the control modules of the system code. As expected, the time-dependent behaviour of the system was driven by physical and chemical changes in the tailings source term over time.

The pyritic, reference tailings were modelled as a layered system; consisting of an upper, unsaturated zone and a lower, saturated zone. The oxidation of pyrite in the unsaturated zone, and the dissolution and reprecipitation of a gypsum-radium solid solution governed the behaviour of numerous other chemical species. Species modelled included radium, thorium, radon, calcium, iron and sulphate as well as pH.

The incorporation of time dependency also gave rise to complications regarding the treatment of parameters which are not only uncertain, but vary from year to year (eg. mean annual windspeed). Several methods for dealing with such parameters were examined and are discussed in this report. The method employed during Phase II involved randomly sampling such parameters only once for each trial, from a distribution adjusted for the length of the simulation period.

The time-dependent component models, and revised control modules, were integrated to form the Uranium Tailings Assessment Program (UTAP). Documentation of this and a user's manual for this program form Volume 3 of this report.

The computer code (the system model) was used to perform probabilistic analyses of a number of different tailings management scenarios. The results of all analyses are summarized in this report.

In addition "importance analysis" designed to identify important pathways and input parameters, was performed for Scenarios 1 and 2. It is a combination of traditional sensitivity and uncertainty analysis, and employed methods found to be most appropriate from auxiliary Phase II studies.

In summary, all of the Phase II objectives have been fulfilled, all investigations were successfully carried out and the results are discussed in the appropriate sections of this report. The report is presented in three volumes, as indicated below, with many of the more technical discussions and details reserved for Volumes 2 and 3.

Volume 1:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 1, Main Report.
Volume 2:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 2, Workshop Notes and Topical Discussion Papers.
Volume 3:	Probabilistic Model Development for the Assessment of Uranium Mill Tailings in Canada - Phase II, Volume 3, Probabilistic Analysis Users Manual and Documentation for the Uranium Tailings Assessment Program (UTAP)

6.2 Recommendations for Further Work

A number of the Phase II investigations have given rise to issues which should be addressed in Phase III studies. Investigation of many of these issues will bring the NUTP closer to achieving a viable, realistic framework for the probabilistic assessment of uranium mill tailings. Further study should also improve the presentation, interpretation and understanding of probabilistic results. The recommendations are presented in point form below:

- Some consideration should be given to predicting the spatial distribution of effects arising from the tailings site.

This might be most easily done by modelling additional receptors to varying locations from the tailings out to a distance of perhaps 50 km for air pathways. The surface water system could be examined by chaining together river/lake modules with different characteristics.

- Component models for additional radionuclides and chemical species should be developed and included in the system model.
- The inclusion of further pathways of exposure (eg. water-terrestrial animal-dose) should be investigated and included in the system code as appropriate.
- Inclusion of further species or pathways will require the estimation of further parameters. All current model parameters should be subjected to ongoing review and revision.
- All current component models, and any additional component models should be reviewed for their ability to incorporate temporal (and spatial) variation. Phase II work indicated that the models must be developed with a specific application in mind. Not all models are appropriate for all applications.
- Thus far, the system modelled has been (relatively) "well-behaved" with time. Consideration should be given to processes which change the tailings over time. Gradual erosion of tailings covers and tailings are of prime interest in this region. In addition, consideration should be given to treatment of some episodic events (such as floods, forest fires, etc) which have the capacity to alter the time-dependent behaviour of the system.
- The Phase II work has given rise to some interesting concepts, such as methods for incorporating parameter correlation and confidence bands on output distributions. Thus far, these have primarily been considered on a theoretical, or prototype level. Effort should be made to

provide a practical approach for efficient implementation of some of the recommended methods/concepts.

- The incorporation of parameter correlation, or parameters which vary in time and space may confound the "importance analysis" techniques applied in Phase II. Continued attention should thus be given to developing appropriate sensitivity and/or uncertainty analysis methods.
- Finally, work should continue on investigating efficient, effective methods of displaying/analysing/interpreting the results of probabilistic analysis. The Phase II studies saw much progress in this area, but more rigorous, acceptable and/or understandable methods may be possible. This is a particularly important consideration with regard to time-dependent behaviour.

REFERENCES

Cairns, J.J. and Fleming, K.N., 1977. "STADIC - A Computer Code for Combining Probability Distributions." Prepared for San Francisco Operations Office, U.S. Energy, Research and Development Administration, Contract EY-76-C-03-0167.

Colorado Department of Health (CDOH), 1984. "Fugitive Partculate Emissions." Interoffice Memo from T. Tistinic, Dated July 2.

Cheng, R.C.H. and Iles, T.C., 1983. "Confidence Bands for Cumulative Distribution Functions of Continuous Random Variables." Technometrics, Vol. 25, No. 1, pp. 77-86.

Downing, D.J., Gardner, R.H., and Hoffman, F.O., 1985. "An Examination of Response-Surface Methodologies for Uncertainty Analysis in Assessment Models." Technometrics, Vol. 27, No. 2, pp. 151-163, May.

Draper, V. and Smith, H., 1981. "Applied Regression Analysis." Second Edition, John Wiley & Sons Inc., New York.

Edwards, A.L., 1979. "Multiple Regression and the Analysis of Variance and Covariance." W.H. Freeman & Co.

Gardner, R.H., 1985. (in press). "Error Analysis and Sensitivity Analysis." Encyclopedia of Systems and Control, M. Singh (ed.), Pergamon Press, London.

Hendrickson, R.G., 1984. "A Survey of Sensitivity Analysis Methodology." U.S. Department of Commerce, National Bureau of Standards, Report NBSIR 84-2814, Washington, DC.

Hoffman, F.O. and Gardner, R.H., 1983. "Evaluation of Uncertainties in Environmental Radiological Assessment Models." In: Radiological Assessment: A Textbook on Environmental Dose Assessment, J.E. Till and H.R. Meyer (eds.), NUREG/CR-3332, ORNL-5968. U.S. Nuclear Regulatory Commission, Washington, D.C.

Iman, R.L. and Conover, W.J., 1982. "A Distribution-free Approach to Inducing Rank Correlation Among Input Variables for Simulation Studies." Technometrics.

Iman, R.L. and Helton, J.C., 1985. "A Comparison of Uncertainty and Sensitivity Analysis Techniques for Computer Models." NUREG/CR-3904; SAND84-1461 Sandia National Laboratories, Albuquerque, New Mexico.

INTERA Environmental Consultants Inc., 1983. "A Proposed Approach to Uncertainty Analysis." Prepared for Office of Nuclear Waste Isolation (ONWI), Battelle Memorial Institute, Columbus, Ohio, ONWI-488.

International Commission on Radiological Protection (ICRP), Publication 23, 1975. "Report of the Task Group on Reference Man." Published by Pergamon Press.

Kendall, M., and Stuart, A., 1961. "The Advanced Theory of Statistics, Vol. 2: Inference and Relationship." 1st Edition, Hafner Publishing Co.

Kleijnen, J.P.C., 1974. "Statistical Techniques in Simulation - Part 1." Marcel Dekker, Inc., New York.

Marse, K., and Roberts, S.D., 1983. "Implementing a Portable Uniform (0,1) Generator." Simulation.

Reilly, P.M., 1985 (April), Personal Communication.

Reilly, P.M. and Johri, H.P., 1969. "Decision Making Through Opinion Analysis." Chemical Engineering, Apr. 7.

Rogers, V.C., Overmeyer, R.F., Putzig, K.M., Jensen, C.M., Nielson, K.K., and Sermon, B.W., 1980. "Characterization of Uranium Tailings Cover Materials for Radon Flux Reduction." Prepared by Ford, Bacon and Davis Utah, Inc., for U.S. Nuclear Regulatory Commission, Report NUREG/CR-1081., FBDU-218-2.

Rose, K.A., and Swartzman, G.L., 1981. "A Review of Parameter Sensitivity Methods Applicable to Ecosystem Models." NUREG/CR-2016, U.S. Nuclear Regulatory Commission, Washington, DC.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada - Phase I." Research Report prepared for the National Uranium Tailings Program, CANMET, EMR, under Supply and Services Canada Contract No. 0SQ83-00236.

Spicer, C.C., 1972. "Algorithm AS52 Calculation of Power Sums of Deviations about the Mean." Applied Statistics 21, 226-227.

Utah State University, 1982. "Probabilistic Modelling of Tailings Embankment Designs." For the U.S. Bureau of Mines (161-82), Washington, D.C., PB83-122598.

Wimmenauer, C., 1971. "Monte-Carlo Simulation: An Overview." For the Quality Evaluation Department; Naval Ammunitions Depot, Crane, Indiana.

APPENDIX A

DETAILED RESULTS OF SENSITIVITY INDEX ANALYSIS AND STEPWISE REGRESSION

A.1 Description of Tables

The sensitivity index tables (Table A.1 and Table A.7) are largely self explanatory. The plus and minus signs under the column entitled "influence" refer to the direction of response of the total dose to a change in magnitude of the parameters listed in the left hand column. A plus sign infers a direct proportionality while a minus sign indicates an inverse proportionality.

The stepwise regression tables (Tables A.2 to A.6 and Table A.8) require explanation. The first two columns in the table contain the regression step number and the name of the variable included in that step. Each step is a separate regression involving the variable named in that step and the variables above it in the table. The remaining columns contain information on that particular regression.

The multiple R and RSQ (R^2) are the overall regression coefficient and the overall regression coefficient squared. The next column is merely the increase in the overall R^2 coefficient due to the inclusion of that variable in the regression. In general the larger the increase R^2 the more important the variable. The overall regression coefficient is not to be confused with the partial correlation coefficient which is the criteria on which a variable is included in the regression (see the article by Beals et al. in Volume 2 for details of stepwise regression). This order of inclusion of variables in stepwise regression was used as the basis for ranking variables.

The F value is the F statistic for that particular regression variable. In general the larger the F statistic the more important the variable. The sum of squares values provide further information on the relative importance of the parameters.

Table A.1

Sensitivity Indices for Total Dose at T=0 and 200 for Scenario 1

	T=0		YEAR	T=200	
VARIABLE	SENSITIVITY INDEX	INFLUENCE	VARIABLE	SENSITIVITY INDEX	INFLUENCE
BF	.9139	+	BF	.9250	+
CONCRA	.6667	+	RAINGP	.8111	+
RAINTH	.5340	-	DISTCO	.7642	+
OF	.4456	+	CONCRA	.6667	+
DISTCO	.4397	+	NPRCP	.5852	+
XK1	.2948	-	FACTCO	.4821	-
EMANAT	.2787	+	OF	.4474	+
RHOTAL	.2765	+	KSUBDR	.4443	+
RAINGP	.2464	+	SULPRD	.2719	+
FISHIN	.2014	+	RHOTAL	.2435	+
FACTCO	.1959	-	KPRIME	.2332	-
MPCH2O	.1656	-	FISHIN	.2236	+
BETA	.1360	+	RAINTH	.2071	-
SULPRD	.1306	-	BETA	.1300	+
F	.0959	+	XK3	.0863	-
O2DIFF	.0847	-	EMANAT	.0826	+
FLOCA	.0734	+	O2DIFF	.0775	-
DEPVEL	.0727	+	XK1	.0662	-
VDISP	.0658	-	MPCH2O	.0447	-
FDEPSF	.0430	+	THFRAC	.0278	+
FACTA	.0289	+	F	.0248	+
KPRIME	.0260	+	THDISS	.0245	-
UWIND	.0235	-	FLOCA	.0234	+
YDEN3	.0231	-	DEPVEL	.0230	+
NPRCP	.0142	-	VDISP	.0185	-
VEGLAM	.0139	-	FDEPSF	.0135	+
FDEPED	.0097	+	UWIND	.0111	-
XK3	.0088	-	FACTA	.0090	+
FWIND2	.0083	+	YDEN3	.0072	-
FWIND1	.0033	+	VEGLAM	.0043	-
FDEPSF	.0022	+	FDEPED	.0030	+
FDEPED	.0022	+	FWIND2	.0023	+
DUREXP	.0016	+	FWIND1	.0010	+
FACTB	.0014	+	FDEPSF	.0007	+
YDEN4	.0012	-	FDEPED	.0007	+
FPREC	.0011	-	DUREXP	.0005	+
VEGLAM	.0007	-	TVOID	.0004	-
HSRC	.0002	-	YDEN4	.0004	-
DUREXP	.0001	+	FACTB	.0004	+
FAUT1	.0001	+	FPREC	.0003	-
FMEAT	.0001	+	VEGLAM	.0002	-
KSUBDR	.0000	+	HSRC	.0000	-
TVOID	.0000	+	FDEPSF	.0000	+
THDISS	.0000	+	FDEPSF	.0000	+
THFRAC	.0000	+	FDEPED	.0000	+
FDEPSF	.0000	+	FDEPED	.0000	+
FDEPSF	.0000	+	DUREXP	.0000	+
FDEPED	.0000	+	DUREXP	.0000	+
FDEPED	.0000	+	DUREXP	.0000	+
DUREXP	.0000	+	VEGLAM	.0000	-
DUREXP	.0000	+	VEGLAM	.0000	+
VEGLAM	.0000	-	YDEN1	.0000	-
VEGLAM	.0000	-	YDEN2	.0000	-
YDEN1	.0000	-	FBI	.0000	+
YDEN2	.0000	-	FAUT1	.0000	+
FBI	.0000	+	FAUT2	.0000	+
FAUT2	.0000	+	FMEAT	.0000	+

Table A.2

Stepwise Regression on Total Dose at T=0 for Scenario 1

STEP NUMBER	VARIABLE		MULTIPLE		INCREASE IN RSQ	F VALUE TO ENTER OR REMOVE	SUM OF SQUARES ADDED
	ENTERED	REMOVED	R	RSQ			
1	MPCH2O		.5414	.2931	.2931	40.6337	1.5149
2	CONCRA		.6813	.4641	.1710	30.9617	.8841
3	EMANAT		.7776	.6047	.1406	34.1370	.7265
4	BF		.8341	.6957	.0910	28.3968	.4702
5	OF		.8767	.7686	.0729	29.6056	.3767
6	F		.9180	.8427	.0741	43.7932	.3830
7	DISTCO		.9305	.8658	.0231	15.8390	.1194
8	RAINTH		.9379	.8796	.0138	10.4267	.0713
9	RHOTAL		.9459	.8948	.0152	12.9880	.0785
10	FISHIN		.9527	.9076	.0129	12.4033	.0665
11	DEPVEL		.9567	.9152	.0076	7.9103	.0394
12	FAUT2		.9601	.9218	.0066	7.3245	.0340
13	FLOCA		.9633	.9279	.0061	7.2291	.0313
14	YDEN4		.9661	.9334	.0055	7.0584	.0286
15	THDISS		.9676	.9362	.0028	3.6649	.0144
16	YDEN1		.9690	.9390	.0028	3.7450	.0142
17	FACTCO		.9702	.9413	.0023	3.2614	.0121
18	FDEPED		.9714	.9436	.0023	3.3221	.0120
19	TVOID		.9725	.9458	.0022	3.2815	.0115
20	DUREXP		.9737	.9481	.0023	3.5131	.0119
21	SULPRD		.9747	.9501	.0019	3.0262	.0100
22	DUREXP		.9756	.9519	.0018	2.8636	.0093
23	FDEPED		.9766	.9537	.0019	3.0851	.0097
24	FBI		.9772	.9550	.0012	2.0589	.0064
25	FAUT1		.9778	.9562	.0012	1.9972	.0061
26	YDEN2		.9784	.9572	.0010	1.7619	.0053
27	FACTA		.9790	.9584	.0012	2.0774	.0062
28	FDEPSF		.9795	.9594	.0010	1.7699	.0052
29	O2DIFF		.9800	.9605	.0011	1.8882	.0055
30	KSUBDR		.9804	.9612	.0007	1.3298	.0039
31	VEGLAM		.9808	.9619	.0007	1.2137	.0035
32	FDEPEP		.9810	.9624	.0005	.9289	.0027
33	VEGLAM		.9813	.9629	.0005	.9211	.0027
34	XK1		.9816	.9636	.0007	1.2003	.0035
35	VEGLAM		.9819	.9640	.0004	.7712	.0022
36	NPRCP		.9821	.9645	.0004	.7943	.0023
37	DUREXP		.9823	.9648	.0003	.5819	.0017
38	RAINGP		.9824	.9651	.0003	.5485	.0016
39	VDISP		.9825	.9654	.0003	.4563	.0014
40	THFRAC		.9827	.9657	.0003	.5317	.0016
41	VEGLAM		.9829	.9661	.0004	.6270	.0019
42	DUREXP		.9831	.9665	.0004	.7066	.0021
43	FWIND1		.9833	.9668	.0003	.5013	.0015
44	FDEPED		.9834	.9670	.0002	.3548	.0011
45	FMEAT		.9835	.9672	.0002	.3622	.0011
46	KPRIME		.9836	.9674	.0002	.3051	.0010
47	FACTB		.9836	.9676	.0001	.2374	.0008
48	FDEPSF		.9837	.9676	.0001	.1366	.0004
49	YDEN3		.9837	.9677	.0001	.1286	.0004
50	FPREC		.9838	.9678	.0001	.1420	.0005
51	XK3		.9838	.9679	.0001	.1126	.0004
52	FDEPSF		.9839	.9680	.0001	.1165	.0004
53	FWIND2		.9839	.9680	.0000	.0497	.0002
54	FDEPSF		.9839	.9680	.0000	.0628	.0002
55	BETA		.9839	.9681	.0000	.0640	.0002
56	HSRC		.9839	.9681	.0000	.0309	.0001

Table A.3

Stepwise Regression on Total Dose at T=200 for Scenario 1

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	MULTIPLE R	MULTIPLE RSQ	INCREASE IN RSQ	F VALUE TO ENTER OR REMOVE	SUM OF SQUARES ADDED
1	KSUBDR		.5342	.2854	.2854	39.1397	2.9757
2	RAINGP		.6172	.3810	.0956	14.9726	.9963
3	DISTCO		.6774	.4589	.0780	13.8347	.8130
4	CONCRA		.7409	.5489	.0900	18.9479	.9381
5	BF		.7810	.6099	.0610	14.6964	.6359
6	MPCH2O		.8071	.6514	.0415	11.0680	.4326
7	RHOTAL		.8226	.6767	.0253	7.1900	.2635
8	OF		.8353	.6977	.0210	6.3300	.2193
9	EMANAT		.8447	.7135	.0158	4.9688	.1649
10	FACTCO		.8521	.7261	.0126	4.0871	.1312
11	RAINTH		.8595	.7388	.0127	4.2869	.1327
12	VEGLAM		.8657	.7494	.0106	3.6835	.1106
13	TVOID		.8701	.7571	.0077	2.7170	.0800
14	THDISS		.8728	.7618	.0047	1.6849	.0492
15	F		.8757	.7669	.0051	1.8261	.0528
16	O2DIFF		.8780	.7709	.0041	1.4688	.0423
17	YDEN3		.8797	.7739	.0030	1.0830	.0311
18	FDEPED		.8817	.7774	.0035	1.2699	.0364
19	YDEN2		.8833	.7802	.0028	1.0058	.0288
20	FPREC		.8846	.7825	.0023	.8486	.0244
21	FISHIN		.8862	.7853	.0028	1.0098	.0290
22	FDEPSF		.8879	.7883	.0030	1.1027	.0316
23	FDEPEP		.8892	.7907	.0023	.8489	.0244
24	DEPVEL		.8909	.7937	.0031	1.1120	.0319
25	YDEN1		.8925	.7966	.0029	1.0632	.0305
26	KPRIME		.8941	.7995	.0028	1.0354	.0297
27	FDEPED		.8953	.8016	.0021	.7615	.0219
28	FDEPSF		.8964	.8035	.0019	.6862	.0198
29	YDEN4		.8973	.8052	.0017	.6198	.0180
30	FMEAT		.8986	.8075	.0023	.8319	.0242
31	DUREXP		.8993	.8088	.0013	.4457	.0131
32	XK3		.9001	.8102	.0014	.5101	.0151
33	FBI		.9008	.8114	.0012	.4235	.0126
34	FDEPSF		.9017	.8131	.0017	.5856	.0176
35	VEGLAM		.9023	.8142	.0011	.3799	.0115
36	HSRC		.9028	.8151	.0009	.2972	.0091
37	FLOCA		.9033	.8160	.0009	.3080	.0095
38	FACTB		.9036	.8165	.0005	.1759	.0055
39	XK1		.9039	.8170	.0004	.1369	.0044
40	DUREXP		.9042	.8175	.0005	.1761	.0057
41	DUREXP		.9044	.8180	.0005	.1488	.0049
42	UWIND		.9046	.8184	.0004	.1323	.0044
43	FWIND1		.9049	.8188	.0004	.1228	.0041
44	VDISP		.9050	.8190	.0002	.0670	.0023
45	SULPRD		.9051	.8193	.0003	.0857	.0030
46	FDEPSF		.9053	.8195	.0002	.0642	.0023
47	FAUT1		.9053	.8197	.0002	.0435	.0016
48	VEGLAM		.9054	.8198	.0002	.0494	.0018
49	DUREXP		.9055	.8200	.0001	.0357	.0013
50	FWIND2		.9056	.8200	.0001	.0178	.0007
51	FAUT2		.9056	.8201	.0001	.0246	.0010
52	THFRAC		.9056	.8202	.0001	.0156	.0006
53	FACTA		.9057	.8202	.0001	.0135	.0006

Table A.4

Stepwise Regression on Water Consumption Dose at T=200 for Scenario 1

STEP NUMBER	VARIABLE		MULTIPLE		INCREASE IN RSQ	F VALUE TO ENTER OR REMOVE	SUM OF SQUARES ADDED
	ENTERED	REMOVED	R	RSQ			
1	KSUBDR		.5598	.3134	.3134	44.7279	5.0065
2	RATWGP		.7047	.4966	.1832	35.2988	2.9268
3	DISTCO		.7863	.6182	.1217	30.5942	1.9437
4	CONCRA		.8253	.6810	.0628	18.7080	1.0034
5	OF		.8417	.7084	.0274	8.8218	.4372
6	RHOTAL		.8529	.7275	.0191	6.5115	.3048
7	FACTCO		.8608	.7409	.0134	4.7625	.2143
8	RAINTH		.8657	.7495	.0086	3.1075	.1367
9	NPRCP		.8692	.7556	.0061	2.2432	.0973
10	XK3		.8715	.7595	.0039	1.4595	.0630
11	SULPRD		.8728	.7618	.0023	.8346	.0361
12	O2DIFF		.8739	.7637	.0019	.7169	.0311
13	THDISS		.8746	.7649	.0012	.4321	.0189
14	THFRAC		.8747	.7650	.0001	.0470	.0021
15	KRPIME		.8747	.7651	.0001	.0453	.0020
16	BETA		.8748	.7652	.0001	.0355	.0016

Table A.5

Stepwise Regression on Radon Inhalation Dose at T=0 for Scenario 1

STEP NUMBER	VARIABLE		MULTIPLE		INCREASE IN RSQ	F VALUE TO ENTER OR REMOVE	SUM OF SQUARES ADDED
	ENTERED	REMOVED	R	RSQ			
1	MPCH2O		.6717	.4511	.4511	80.5533	6.8321
2	EMANAT		.9197	.8459	.3947	248.4703	5.9780
3	CONCRA		.9429	.8890	.0431	37.3053	.6531
4	F		.9620	.9255	.0365	46.4953	.5523
5	OF		.9841	.9684	.0429	127.5417	.6496
6	RHOTAL		.9885	.9771	.0087	35.2283	.1315
7	VDISP		.9900	.9801	.0031	14.2324	.0465
8	FUSWD2		.9906	.9812	.0011	5.1126	.0160
9	UWIND		.9909	.9819	.0007	3.5033	.0107
10	RAINTH		.9911	.9822	.0003	1.6807	.0051
11	KPRIME		.9912	.9825	.0003	1.5221	.0046
12	THFRAC		.9913	.9827	.0002	.9112	.0027
13	MPRCP		.9914	.9829	.0001	.6700	.0020
14	SULPRD		.9914	.9830	.0001	.4800	.0015
15	HSRC		.9915	.9830	.0001	.4769	.0015
16	RAINGP		.9915	.9831	.0001	.3985	.0012
17	FACTCO		.9915	.9831	.0000	.0171	.0001
18	DISTCO		.9915	.9831	.0000	.0110	.0000

Table A.6

Summary Table for Stepwise Regression on pH at T=0 for Scenario 1

STEP NUMBER	VARIABLE		MULTIPLE		INCREASE IN RSQ	F VALUE TO ENTER OR REMOVE	SUM OF SQUARES ADDED
	ENTERED	REMOVED	R	RSQ			
1	SULPRD 2		.7707	.5940	.5940	143.3626	.0500
2	O2DIFF 4		.9560	.9140	.3200	360.9610	.0270
3	RHOTAL 3		.9666	.9342	.0202	29.5203	.0017
4	KPRIME 5		.9710	.9428	.0085	14.1669	.0007
5	FACTCO 7		.9731	.9468	.0041	7.1925	.0003
6	NPRCP 6		.9742	.9490	.0022	3.9665	.0002

Table A.7

Sensitivity Indices at T=100 for Scenario 2

VARIABLE	SENSITIVITY INDEX (INFLUENCE)
KSUBDR	.9691 (-)
BF	.9278 (+)
RAINGP	.8896 (+)
CONCRA	.6665 (+)
SULPRD	.5968 (+)
NPRCP	.5732 (+)
FACTCO	.4585 (-)
OF	.4487 (+)
O2DIFF	.4415 (+)
RHOTAL	.3138 (+)
DISTCO	.2781 (+)
FISHIN	.2299 (+)
BETA	.1775 (+)
KPRIME	.1603 (-)
XK1	.1117 (-)
RAINTH	.0695 (-)
XK3	.0532 (-)
EMANAT	.0205 (+)
MPCH2O	.0108 (-)
THDISS	.0075 (-)
F	.0059 (+)
FLOCA	.0056 (+)
DEPVEL	.0055 (+)
THFRAC	.0044 (+)
VDISP	.0044 (-)
UWIND	.0033 (-)
FDEPSF	.0032 (+)
FACTA	.0021 (+)
YDEN3	.0017 (-)
VEGLAM	.0010 (-)
TVOID	.0007 (-)
FDEPED	.0007 (+)
FWIND1	.0005 (+)
FWIND2	.0005 (+)
FDEPSF	.0002 (+)
FDEPED	.0002 (+)
FPREC	.0001 (-)
DUREXP	.0001 (+)
VEGLAM	.0001 (-)
YDEN4	.0001 (-)
FACTB	.0001 (+)
HSRC	.0000 (-)
FDEPSF	.0000 (+)
FDEPSF	.0000 (+)
FDEPED	.0000 (+)
FDEPED	.0000 (+)
DUREXP	.0000 (+)
DUREXP	.0000 (+)
DUREXP	.0000 (+)
VEGLAM	.0000 (-)
VEGLAM	.0000 (+)
YDEN1	.0000 (-)
YDEN2	.0000 (+)
FBI	.0000 (+)
FAUT1	.0000 (+)
FAUT2	.0000 (+)
FMEAT	.0000 (+)

Table A.8

Stepwise Regression on Log Total Dose at T=100 for Scenario 2

STEP	VARIABLE		MULTIPLE		INCREASE	F VALUE TO	SUM OF SQUARES
NUMBER	ENTERED	REMOVED	R	RSQ	IN RSQ	ENTER OR REMOVE	ADDED
1	KSUBDR		.8052	.6483	.6483	180.6319	10.4185
2	BF		.8495	.7217	.0734	25.5938	1.1801
3	CONCRA		.8843	.7820	.0603	26.5459	.9688
4	RAINGP		.9179	.8425	.0605	36.4593	.9717
5	OF		.9356	.8753	.0328	24.7285	.5273
6	MPCH2O		.9407	.8849	.0096	7.7969	.1551
7	TVOID		.9436	.8904	.0054	4.5672	.0875
8	UWIND		.9457	.8944	.0040	3.4797	.0649
9	VEGLAM		.9474	.8976	.0032	2.8549	.0522
10	VEGLAM		.9496	.9017	.0041	3.6736	.0652
11	HSRC		.9513	.9050	.0033	3.0933	.0536
12	FACTCO		.9525	.9072	.0022	2.0531	.0352
13	DISTCO		.9535	.9091	.0019	1.7964	.0305
14	FISHIN		.9545	.9111	.0019	1.8615	.0313
15	KPRIME		.9555	.9130	.0020	1.8834	.0313
16	FBI		.9564	.9146	.0016	1.5574	.0257
17	FWIND1		.9570	.9159	.0013	1.2622	.0208
18	VDISP		.9577	.9172	.0013	1.2791	.0210
19	F		.9584	.9185	.0012	1.2097	.0198
20	O2DIFF		.9590	.9197	.0013	1.2685	.0207
21	THFRAC		.9598	.9212	.0015	1.4497	.0235
22	RHOTAL		.9606	.9227	.0015	1.5130	.0244
23	YDEN1		.9614	.9243	.0016	1.6201	.0259
24		UWIND	.9614	.9243	.0000	.0043	-.0001
25	XK3		.9619	.9252	.0008	.8470	.0134
26	DEPVEL		.9622	.9259	.0007	.7172	.0114
27	VEGLAM		.9626	.9266	.0007	.7179	.0114
28	FDEPED		.9629	.9272	.0006	.6101	.0098
29	YDEN4		.9633	.9279	.0007	.7345	.0118
30	EMANAT		.9636	.9285	.0006	.5877	.0095
31	YDEN2		.9639	.9292	.0006	.6201	.0101
32	YDEN3		.9643	.9299	.0007	.7038	.0115
33	FDEPEP		.9646	.9305	.0006	.5845	.0096
34	FDEPSF		.9648	.9309	.0004	.3980	.0066
35	RAINTH		.9652	.9316	.0007	.6651	.0111
36	FDEPED		.9655	.9322	.0007	.6239	.0105
37	FAUT1		.9660	.9331	.0009	.8754	.0147
38	SULPRD		.9663	.9338	.0006	.6009	.0102
39	UWIND		.9666	.9344	.0006	.5691	.0097
40	FDEPSF		.9669	.9349	.0005	.4800	.0082
41	XK1		.9671	.9352	.0003	.2963	.0051
42	DUREXP		.9672	.9356	.0004	.3274	.0057
43	FMEAT		.9674	.9358	.0002	.1871	.0033
44	FDEPED		.9674	.9359	.0002	.1466	.0026
45	THDISS		.9675	.9361	.0002	.1371	.0025
46	NPRCP		.9676	.9363	.0002	.1550	.0029
47	FDEPSF		.9677	.9364	.0001	.0870	.0016
48	FACTA		.9677	.9365	.0001	.0785	.0015
49	FWIND2		.9677	.9365	.0001	.0573	.0011
50		XK3	.9677	.9365	.0000	.0028	-.0001
51	DUREXP		.9678	.9366	.0000	.0353	.0007
52	DUREXP		.9678	.9366	.0001	.0460	.0009
53	DUREXP		.9678	.9367	.0000	.0267	.0005
54	BETA		.9678	.9367	.0000	.0266	.0006
55	FAUT2		.9678	.9367	.0000	.0147	.0003
56	FACTB		.9679	.9367	.0000	.0120	.0003
57	VEGLAM		.9679	.9368	.0000	.0156	.0003

APPENDIX B

DETAILED LISTING OF INPUT PARAMETERS

KEY TO TABLE B.1

Distribution Type		Distribution Specifications
C	Constant	single numerical value
LN	Log Normal	geometric mean, geometric standard deviation, minimum, maximum
N	Normal	arithmetic mean, arithmetic standard deviation, minimum, maximum
T	Triangular	minimum, mode (peak), maximum
U	Uniform	minimum, maximum
LU	Log Uniform	minimum, maximum

Table B.1

MODEL PARAMETERS FOR SCENARIOS 1 AND 2

Computer Symbol	Subroutine	Definition (Units)	Distribution
BETA	LAKE	parameter relating the rate constant k_2 to rate constant k_1	LN (0.18, 1.35, 0.1, 0.33)
BF	FISH	water to fish flesh transfer coefficient ($Bq.g^{-1}$ (wet wt).$Bq^{-1}.m^3$)	LN ($1.6x10^{-5}$, 3.5, $1.0x10^{-6}$, $6.0x10^{-4}$)
CONCRA	SOURCE	initial solid phase activity of Ra-226 in tailings ($Bq.g^{-1}$)	T(5.0, 11.0, 15.0)
DEPVEL	VEGDEP	deposition velocity ($m.s^{-1}$)	U(0.003, 0.03)
DISTCO	SOURCE	radium solid to liquid distribution coefficient (unsaturated zone)	U(.01, 0.1)
DUREXP(J)	VEGDEP	duration of exposure to deposition (s)	see below
DUREXP(1)	VEGDEP	lichens	U($90x10^6$, $110x10^6$)
DUREXP(2)	VEGDEP	pasture grass	U($2.7x10^6$, $3.1x10^6$)
DUREXP(3)	VEGDEP	above ground vegetation	U($4.7x10^6$, $5.6x10^6$)
DUREXP(4)	VEGDEP	below ground vegetation	U($4.7x10^6$, $5.6x10^6$)
EMANAT	RADON	radon emanation coefficient of tailings	T(0.02, 0.2, 0.9)
F	DOSEIN	equilibrium factor for radon daughters to radon	T(0.2, 0.4, 0.7)
FACTA	DOSEIN	fraction of activity in above ground vegetation after food preparation	U(0.5, 1.0)
FACTB	DOSEIN	fraction of activity in below ground vegetation after food preparation	U(0.5, 1.0)
FACTCO	SOURCE	activity coefficient (for divalent ions)	T(0.2, 0.3, 0.5)
FAUT(J)	VEGEAT	fraction of annual intake that is local vegetation	see below
FAUT(1)	VEGEAT	lichens	T(0.05, 0.2, 0.5)
FAUT(2)	VEGEAT	pasture grass	T(0.05, 0.2, 0.5)
FBI	VEGEAT	feed-to-meat transfer coefficient for caribou ($Bq.g^{-1}$ (wet wt).$Bq^{-1}.yr$)	U($0.3x10^{-9}$, $2.5x10^{-9}$)
FDEPED(J)	VEGDEP	fraction of plant deposition retained on edible portions	see below
FDEPED(1)	VEGDEP	lichens	U(0.8, 1.0)
FDEPED(2)	VEGDEP	pasture grass	U (0.8, 1.0)
FDEPED(3)	VEGDEP	above ground vegetation	U (0.8, 1.0)
FDEPED(4)	VEGDEP	below ground vegetation	U(0.05, 0.15)
FDEPSF(J)	VEGDEP	fraction of total deposition retained on vegetation	see below
FDEPSF(1)	VEGDEP	lichens	U (0.8, 1.0)
FDEPSF(2)	VEGDEP	pasture grass	N(0.47, 0.3, 0.01, 1.0)
FDEPSF(3)	VEGDEP	above ground vegetation	U (0.1, 0.3)
FDEPSF(4)	VEGDEP	below ground vegetation	U (0.1, 0.3)
FISHIN	DOSEIN	average annual intake of fish ($g.yr^{-1}$)	U ($4x10^3$, $12x10^3$)
FLOCA	DOSEIN	fraction of above ground vegetables locally grown	T(0.1, 0.3, 0.6)
FLOCB	DOSEIN	fraction of below ground vegetables locally grown	as per FLOCA
FMEAT	DOSEIN	fraction of annual caribou meat intake that is local	U(0.0, 1.0)

Table B.1 continued

Computer Symbol	Subroutine	Definition (Units)	Distribution
FPREC	DUST	annual number of days with more than 0.25 mm of precipitation (days)	N (123, 18, 84, 156)
FRACGP	SOURCE	mass fraction of gypsum in tailings	U(0.05, 0.06)
FWAT	DOSEIN	fraction of annual water consumption that is local	as per OF
FWIND1	DUST	percentage of time that the wind speed exceeds 5.1 $m.s^{-1}$ at the mean tailings height.	N(35, 6, 23, 46)
FWIND2	AIRDIS	fraction of time wind blows from the source to the receptor	N (0.1, 0.02, 0.06, 0.14)
HSRC	AIRDIS	height of the source (m)	U (1,3)
KPRIME	SOURCE	ratio of O_2 half saturation constant KO_2 to the oxygen concentration at the tailings surface, $[O_2]$ ($mol.m^{-3}.mol^{-1}.m^3$)	LN (0.05, 1.7, 0.02, 0.25)
KSUBDR	SOURCE	Ra-226 liquid to solid transfer coefficient (mixing zone) ($m^3.g^{-1}$)	LN ($80x10^{-6}$, 3.5, $1.9x10^{-6}$, $3.4x10^{-3}$)
MPCH2O	RADON	percent moisture content by weight of tailings	T(5,15,25)
NPRCP	SOURCE	net precipitation ($m.yr^{-1}$)	N(0.4, 0.1, 0.1, 0.7)
OF	DOSEIN	fraction of year at reference site	T (0.4, 0.6, 0.9)
O2DIFF	SOURCE	oxygen diffusivity through tailings ($m^2.yr^{-1}$)	T (30, 70, 130)
RAINGP	SOURCE	fraction of total Ra-226 associated with gypsum	T (0.05, 0.2, 0.6)
RAINTH	SOURCE	mol fraction of Ra-226 ingrowth to solid Th-230 Th-230 matrix which is insoluble	T (0.85, 0.9, 0.95)
RHOTAL	SOURCE	tailings bulk density ($g.m^{-3}$)	N($1.4x10^6$, $0.1x10^6$, $1.1x10^6$, $1.7x10^6$)
SULPRD	SOURCE	sulphate production rate ($mol.g^{-1}.yr^{-1}$)	LN ($1.4x10^{-6}$, 1.5, $0.4x10^{-6}$, $4.2x10^{-6}$)
THDISS	SOURCE	rate constant for amorphous Th dissolution	U(0.01,0.03)
THFRAC	SOURCE	fraction of total Th which is assumed to be immobile	T(0.2, 0.3, 0.4)
TVOID	SOURCE RADON	pore volume fraction	T(0.35, 0.4, 0.45)
UWIND	AIRDIS	mean annual wind speed ($m.s^{-1}$)	N(5.3, 0.5, 3.8, 6.8)
VEGLAM(J)	VEGDEP	decay constant accounting for weathering loss (s^{-1})	see below
VEGLAM(1)	VEGDEP	lichens	U($2x10^{-9}$, $4x10^{-9}$)
VEGLAM(2)	VEGDEP	pasture grass	U($4x10^{-7}$, $6x10^{-7}$)
VEGLAM(3)	VEGDEP	above ground vegetables	U($4x10^{-7}$, $6x10^{-7}$)
VEGLAM(4)	VEGDEP	below ground vegetables	U($4x10^{-7}$, $6x10^{-7}$)
VDISP	AIRDIS	vertical dispersion coefficient (m)	LN(60, 1.08, 47.6, 75.7)
XK1	LAKE	rate of radium removal from the lake water column to lake sediment (yr^{-1})	LN(0.43, 1.85, 0.13, 1.47)
XK3	LAKE	rate of radium removal (burial) to deep lake sediment (yr^{-1})	LN(0.022,1.2, 0.015, 0.03)
YDEN(J)	VEGDEP	yield density ($g.m^{-2}$)	see below
YDEN(1)	VEGDEP	lichens	U (400,600)
YDEN(2)	VEGDEP	pasture grass	U (500,1000)
YDEN(3)	VEGDEP	above ground vegetation	U (1500,2500)
YDEN(4)	VEGDEP	below ground vegetation	U (1500,2500)

Volume

2

NATIONAL
URANIUM
TAILINGS
PROGRAM

Canada

PROBABILISTIC MODEL DEVELOPMENT FOR THE ASSESSMENT OF THE LONG-TERM EFFECTS OF URANIUM MILL TAILINGS IN CANADA - PHASE II

VOLUME 2 - WORKSHOP NOTES AND TOPICAL DISCUSSION PAPERS

DSS File #26SQ.23241-4-1693
Contract # OSQ84-00207

A Research Report Prepared for:

The National Uranium Tailings Program
Canada Centre for Mineral and Energy Technology
Energy, Mines and Resources Canada
555 Booth Street
Ottawa, Ontario
K1A 0G1

by

SENES Consultants Limited
499 McNicoll Avenue
Willowdale, Ontario
M2H 2C9

October 1985

ABSTRACT

This study was conducted to extend and refine the probabilistic assessment tool, developed in a previous study for assessing the long-term effects of uranium mill tailings disposal. The incorporation of time-dependent behaviour was a particularly important aspect of the study. Numerous technical issues relevant to performing or interpreting a probabilistic assessment were examined. A computer code was developed and used to perform probabilistic analyses on several uranium tailings management scenarios.

The study is described in three volumes. This volume, Volume 2, contains the notes of a project workshop and copies of the technical discussion papers prepared by various study participants.

NOTICE

This report was prepared by SENES Consultants Limited (SENES) under contract to the National Uranium Tailings Program (NUTP). Neither the NUTP nor SENES makes any warranty, expressed or implied, or assumes any liability with respect to any damage or loss incurred as a result of the use made of the information contained in this publication.

TABLES OF CONTENTS

Page

1.0 INTRODUCTION 1-1

2.0 SUMMARY NOTES ON PHASE II PROBABILISTIC ASSESSMENT WORKSHOP 2-1
- 2.1 Workshop Objectives (SENES) 2-1
- 2.2 NUTP Update (W. Harrison) 2-1
- 2.3 Update on Phase II Modelling Effort (SENES) 2-2
  - 2.3.1 Phase II Modelling 2-2
  - 2.3.2 Source Term 2-2
  - 2.3.3 Results 2-3
  - 2.3.4 System Code 2-4
- 2.4 Endpoints of Analysis (R. Osborne) 2-4
- 2.5 Treatment of Variable and Uncertain Parameters (S. Kaplan) 2-6
- 2.6 Stepping Through Time (D. Chambers) 2-8
- 2.7 Treatment of Correlation 2-8
  - 2.7.1 Correlated Parameters (T. Andres) 2-8
  - 2.7.2 Estimating Correlation (P. Reilly) 2-9
- 2.8 Statement of Confidence 2-9
  - 2.8.1 Confidence Bands for Probabilistic Density Functions (T. Duever & P. Reilly) 2-9
  - 2.8.2 Confidence Intervals (S. Kaplan) 2-10
- 2.9 Predicting the Dynamic Behaviour of Environmental Systems 2-10
  - 2.9.1 Aquatic Model (R. Gardner, G. Blaylock, O. Hoffman) 2-10
  - 2.9.2 Terrestrial Model (O. Hoffman, R. Gardner) 2-11
- 2.10 Overall Comments 2-11

3.0 SUPPLEMENTARY NOTES ON ENDPOINT FOR ASSESSMENT BY RICHARD V. OSBORNE, AECL, CRNL. 3-1
  3.1 Mandatory Criteria 3-1
  3.2 Preference Criteria 3-2
  3.3 Choice of Parameters 3-2
  3.4 Individual Risk Estimates 3-3
  3.5 Comparison of Tailings Management Options 3-4

APPENDICES: DISCUSSION PAPERS

A THE INTERPRETATION OF A PROBABILISTIC ASSESSMENT OF THE ENVIRONMENTAL EFFECTS OF URANIUM MILL TAILINGS BY S. KAPLAN, L.M. LOWE, AND D.B. CHAMBERS.

B PROBABILITIES OF INDIVIDUAL CELLS OR TAIL AREA UNDER MONTE CARLO SAMPLING BY P.M. REILLY

C CONFIDENCE REGIONS FOR OUTPUT DISTRIBUTIONS FROM MONTE CARLO CALCULATIONS BY P.M. REILLY AND T.A. DUEVER

D SENSITIVITY/UNCERTAINTY ANALYSIS AND EVALUATION OF TIME RESPONSE OF A MODEL OF RADIONUCLIDE ACCUMULATION IN SOIL BY D.I. BEALS, D.B. CHAMBERS AND B.E. HALBERT

E TECHNIQUES FOR HANDLING CORRELATED PARAMETERS BY T. ANDRES AND B. GOODWIN

F MODELLING RADIUM EXCHANGE BETWEEN THE WATER COLUMN AND LAKE SEDIMENT BY J.M. SCHARER AND B.E. HALBERT

G A DYNAMIC MODEL OF RADIUM-226 BEHAVIOUR IN AN OLIGOTROPHIC LAKE BY R.H. GARDNER, B.G. BLAYLOCK AND F.O. HOFFMAN

H THE DATA BASE AVAILABLE FOR ESTIMATING RADIUM-226 BIOACCUMULATION IN FISH FLESH BY B.E. HALBERT, D. STONE AND V.J. CASSADAY.

I THE DATA BASE AVAILABLE TO DEVELOP PARAMETER DISTRIBUTIONS FOR USE IN MODELLING THE TRANSFER OF AIRBORNE RADIOACTIVITY TO TERRESTRIAL VEGETATION BY V.J. CASSADAY, B.E. HALBERT AND D. STONE

J THE TRANSFER OF RADIUM FROM THE ATMOSPHERE TO TERRESTRIAL VEGETATION BY F.O. HOFFMAN, R.H. GARDNER AND B.G. BLAYLOCK

1.0 INTRODUCTION

In 1982 Energy, Mines and Resources Canada established the National Uranium Tailings Program (NUTP) which is intended to pull together previous work on Canadian uranium mill tailings, to fund additional work where significant information gaps are found and to develop the capability to predict the long-term effects of tailings. Such evaluations require the use of mathematical models to simulate the behaviour of the real system. In this regard, SENES Consultants Limited carried out a Phase I study for the NUTP to demonstrate how probabilistic techniques could be used to take account of the natural variability and uncertainty in model parameters in the simulation analysis (SENES 1984).

Subsequently, SENES undertook a Phase II study to extend and refine the probabilistic tools developed in the Phase I study and to investigate various issues of technical interest involved in carrying out and interpreting a probabilistic analysis. The SENES Phase II study is presented in a three-volume series of reports. This report, Volume 2, contains discussion papers prepared on several specific technical issues.

As observed in the Phase I project, the preparation of discussion papers proved most helpful in generating a constructive exchange of technical information amongst the study team participants. Most importantly the studies involved in preparing the discussion papers often involved examination of new topics which had not (at least to our knowledge) been discussed elsewhere. In addition, some of the studies included collection and analysis of data which had previously been dispersed in the literature and not readily available, in a useful form, to the modelling practitioner.

Based in part on the findings of the Phase I study and through dialogue with the Scientific Authority, the study team identified those topics which were considered most appropriate for investigation in Phase II. Subsequently, during late 1984 and early 1985 draft discussion papers were prepared by various participants and distributed among the study team for peer review. In April 1985, a project workshop was held in Toronto to discuss progress on the Phase II investigations and to review specific draft discussion papers. Chapter 2 of this volume presents a summary of the workshop discussions.

Following the April workshop the final version of the technical discussion papers was prepared. Copies of these papers are provided in Appendices A through J of this volume.

The reader is cautioned that several of the papers detail new work and that while every reasonable effort has been made to ensure that the papers are technically accurate and complete, the occasional technical point may have been inadvertently overlooked by the authors and reviewers. SENES and the NUTP would appreciate being advised of any such items which a reader believes he has identified.

2.0 SUMMARY NOTES ON PHASE II PROBABILISTIC ASSESSMENT WORKSHOP

On April 19, 1985, a project workshop was held at the Ramada Inn (Don Valley) in Toronto, Ontario. Workshop attendees included many knowledgeable individuals (see Table 2.1) with a wide range of scientific backgrounds which generated much enthusiastic discussion. The key features of these discussions are summarized in the following notes.

2.1 Workshop Objectives (SENES)

The objectives of this workshop were identified as:

1. Focus the discussion on concepts rather than details.

2. Discuss new subject matter as well as investigations carried out to extend the work done in Phase I.

3. Identify points of agreement and disagreement.

2.2 NUTP Update (W. Harrison)

Two field studies are being conducted in Ontario and Northern Saskatchewan to provide data for demonstrating the techniques developed. The objective of Phase I was to give a quick demonstration of probabilistic modelling techniques using a simplified, one contaminant model. In Phase II, the problem of time-varying parameters and multicontaminant models was to be investigated. To the extent possible, component models to be used were to be off-the-shelf items with the emphasis of the work being on what can be done with them. In Phase III then, better component models will be developed.

2.3 Update on Phase II Modelling Effort (SENES)

The objectives of Phase II have been to:

a) include time-dependent behaviour
b) include additional radionuclides
c) consider non-radioactive contaminants
d) develop transfer factor parameter distributions
e) consider conventional disposal technologies
f) make a preliminary assessment of a generic site
g) investigate and demonstrate sensitivity/uncertainty analysis techniques
h) continue development of the system code

2.3.1 Phase II Modelling

The concept of how the source term is modelled in Phase II was outlined. The tailings mass has been conceptualized in four layers to account for time-varying inputs to the water and air pathways. Auxiliary studies are being undertaken to handle special problems such as immersion and leaching of radionuclides deposited on the soil from the air. Also, the problems of model aggregation are being examined.

2.3.2 Source Term

The basic concept for the source term model is an unsaturated zone in which pyrite oxidation occurs, and a saturated zone which offers a reducing environment. In addition, a mixing zone is included in the top metre of the saturated zone to mimic natural retardation of chemical species movement out of larger tailings piles. Gypsum, is being removed progressively from the top layer (i.e. the unsaturated zone) resulting in a moving gypsum front. The unsaturated zone thus comprises a gypsum-depleted layer and a gypsum-enriched layer. This results in a 4-layer description of the tailings:

a) gypsum-depleted zone
b) gypsum-enriched zone
c) mixing zone
d) reducing zone

Approximately 10% of the precipitation is assumed to enter the groundwater system below the tailings pile, while 90% emerges as surface runoff. In addition to modelling Ra-226, thorium as well as non-radioactive constituents, such as calcium, sulphate and iron, are also considered.

2.3.3 Results

The first set of results shown, depicted the very gradual changes of radium in the tailings solids in the gypsum depleted layer at the top of the tailings pile. The simulation was run for a 500-year period using 5-year time steps. The data were plotted in 50-year increments, with each data point being the mean value of 50 trials. Also shown on the plots were 95% confidence limits on the data plotted at the 50-year time increments. The 95% confidence limits were approximated by multiplying and dividing the geometric mean at each 50-year increment by the corresponding geometric standard deviation squared, to obtain the upper and lower limits respectively. These limits were calculated based on the assumption that the output distribution was lognormal.

Results were also shown for the radium concentration in the mixing zone. The curve was noted to rise to a maximum, which was attributed to the depletion of gypsum out of the top layer, and to subsequently decline. The pH of the porewater in the mixing zone showed an initial increase from t=0 to t=300 years followed by a levelling off.

Q: What happens if the water table moves?

The model can handle problems of this nature via respecifying the depth of the unsaturated zone. To handle catastrophic events such as dam failure and subsequent movement of tailings out of the tailings area, however, may require restructuring of certain parts of the source term model.

2.3.4 System Code

The Parameter Allocation and Control routine (PAC) is of interest. It transfers the inputs to the various simulation models. Different models and additional pathways can be accommodated quite easily in this way by simply including as many simulation routines as are necessary. PAC also allows for modelling time-varying parameters. Random sampling of parameter values for input to each trial is carried out in the routine RANSIM.

2.4 Endpoints of Analysis (R. Osborne)

The reason for going through this modelling effort is to
a) protect individuals
b) choose between options

This procedure leads to the concepts of mandatory and preferential criteria which are discussed in more detail in Section 3.0.

Mandatory criteria are needed to assess the maximum risk to the most-highly exposed individuals in the future. For a member of the critical group, the assurance that the risk is below some limit is required. Another mandatory criterion may require that the frequency of intrusion is low. An additional area of concern may be the maximum concentration of chemicals in public water, food and air. Another type of requirement may be that the average dose in a region be less than a given limit.

Q: How does one define average?

This is, of course, very difficult to define and depends on the context.

A similar problem exists in defining maximum risk. Is it maximum risk for a single year, or over a lifetime (say 50 years)?

The components of risk (R) can be defined as

$$R_{TOTAL} = \underset{\substack{\uparrow \\ \text{normal} \\ \text{risk}}}{R_N} + \sum_{\substack{\text{events} \\ \uparrow \\ \text{events such as} \\ \text{floods, droughts, etc.}}} R_i$$

Q. Should critical group variables be distributed?

One way in which this problem can be partially answered is to do a sensitivity analysis and to assign single constant values to those variables which are found to be unimportant.

It is important to decide, what is meant by critical group. If it is a reference group, then perhaps it is not wise to distribute. Osborne feels it is best to distribute as few parameters as possible.

By preference criteria, we are addressing the question, what is reasonable? Preference will depend on many factors dealing with:

a) populations
b) individuals
c) engineering and costs

Q. What guidelines does R71, the official AECB statement on concepts of nuclear waste disposal, give?

The guidelines given are quite general and the limit below which, say, dose to man must fall are not explicitly stated.

Q. How severe should the requirements for a reference individual be?

For demonstration purposes, the actual characteristics of the reference receptor are not important but their interpretation is. In this regard, it is preferable to consider an average receptor rather than a maximum receptor. In doing so, one hopefully avoids unnecessarily alarming individuals who do not look at the entire report.

Q. Can we afford to have mandatory regulation regardless of cost?

This can lead us into a corner where a new facility will never be approved.

2.5 Treatment of Variable and Uncertain Parameters (S. Kaplan)

The idea was proposed, that the outputs of a Monte Carlo model can have three interpretations depending upon the interpretation of the input parameter distributions:

1. If all inputs represent year-to-year fluctuations or variability, then the outputs represent variability.

2. If the input distributions represent long-term averages and hence represent uncertainty about these long-term values, then the outputs must also be interpreted as representing uncertainty.

3. If the inputs are a mixture of uncertainty and variability then the outputs are not interpretable.

Kaplan's Interpretation 3 gave rise to some disagreement. Reilly pointed out that the output distribution of a Monte Carlo calculation is always interpretable as a subjective probability i.e. it shows what you are entitled to believe about a particular event regardless of the input. Reilly noted that the output distribution can be qualified by considering what the input distributions represent but the interpretation of the outputs is always the same (i.e. it can be interpreted always as a subjective probability) and in each case has a very well defined meaning. In each case the meaning (i.e. interpretation) depends on the nature of the input data, particularly the time scale of time-variant parameters (eg. yearly variability, long-term mean).

Kaplan noted that to be consistent, one must make the input distributions consistent with the question being asked of the model.

The concept of basing a regulatory decision on a limit or even an area is unacceptable since the interpretation of limits or area may change considerably depending on the shape of the probability density function. To make a decision, therefore, the whole curve must be taken into account.

Furthermore, Kaplan argued that the tails of the output distributions are non-robust and have a large amount of uncertainty associated with them because they are derived from the tails of the input distributions and because they have percentage-wise a large amount of shimmer. It was argued that the latter statement is only partially true since tails of the output distribution can be generated from samples drawn from the central part of the input distributions. In other words, in a model with several input parameters, it is not necessary to sample extreme values for the inputs to obtain extreme values of the output. (Subsequent examination of this issue by SENES using a simple additive model consisting of ten independent normal distributions leads to the conclusion that to obtain values in the tails of ouput distributions (say in the upper 0.5% of the output distribution) does not require sampling from the tails of all the input distributions. Our conclusion was that the strong statement that to obtain values in the tails of the output distribution requires all input values to be derived from tails of the input distributions does not alway hold.)

2.6 Stepping Through Time (D. Chambers)

Several approaches to stepping through time were presented and illustrated using results from a soil-leaching model. Typically parameters for each trial are first selected from appropriate distributions and the system is allowed to evolve in time. In the next trial new parameter values are selected.

A difficulty arises when one or more of the input parameters are not only uncertain but vary from year to year. If an annual time step is used, then the value of such a variable can be obtained directly from the distribution of 'annual' values. If a time step of N years is used then the value at each time step is drawn from a distribution with the same mean but whose standard error is smaller by a factor of $1/\sqrt{N}$.

2.7 Treatment of Correlation

2.7.1 Correlated Parameters (T. Andres)

When modelling a system, the real world is divided into two basic types of parameters, independent parameters and dependent parameters. The latter type are derived from the former through some mathematical model. In addition to deciding which parameters are basic, correlation amongst parameters must also be taken into account. Ignoring correlations, leads to changes in the mean and variance of the result and can generate infeasible scenarios and may allow the models to miss important relationships.

Correlation among parameters may be deduced from research programs or may be elicited from expert opinion. These methods may not prove very satisfactory for ascertaining multivariate correlations. Three methods which are proposed for enforcing correlations in the simulations are Kleijnen's correlated generation technique, Iman and Conover's rank correlation and construction of functional relationships. Of these three the construction of functional relationships, when possible, is the preferred method since it:

a) is explicit
b) results in a model that can be tested and refined
c) exposes the true number of parameters in a model
d) encourages the use of nonlinear relationships where appropriate

Andre's preferred method then is to construct functional relationships and to continue the modelling process until the remaining independent variables are uncorrelated.

2.7.2 Estimating Correlation (P. Reilly)

A method for estimating the correlation among parameters given the coefficient of correlation was discussed, which uses a linear programming approach.

2.8 Statement of Confidence

2.8.1 Confidence Bands for Probability Density Functions (T. Duever and P. Reilly)

A method was presented for generating confidence bands for Monte Carlo outputs based on the method by Cheng and Iles. The method consists of:

1. Fitting a probability density function (pdf) of known form to the observed frequency data.
2. Generating a joint confidence region for the estimated parameters.
3. Calculating the confidence bands by substituting into the known form of the pdf all values of the parameters in the joint confidence region.

A model for the observed frequency data was proposed which is flexible and ensures positive probability values every time.

The recommendation was made that this method should be applied to the cumulative distribution function.

2.8.2 Confidence Intervals (S. Kaplan)

Confidence intervals for individual points on the pdf are calculated using a Bayesian approach. A numerical example was shown illustrating how our knowledge about an event changes as the evidence accumulates.

There is some question about which prior should be used for this method, however, it was pointed out that after relatively few Monte Carlo trials, the influence of the prior on the output is negligible.

2.9 Predicting the Dynamic Behaviour of Environmental Systems

2.9.1 Aquatic Model (R. Gardner, G. Blaylock, O. Hoffman)

The model consists of donor controlled linear compartments. This model considers annual dynamics only and does not take seasonal variation into account.

The information required for each compartment is the mass amount in the compartment as well as the turnover time. Interestingly, it was noted that there is very little data available on which to build the model.

Initially the results of the aquatic model showed some serious internal inconsistencies. To remedy this, a deterministic calculation was performed to check the model and then the uncertainties were added.

One problem encountered in applying the data in this model, is that all the data is collected in the summer and may not apply to the winter season for example. It was pointed out that the quality of data is so poor, that the effect of statistical problems like "shimmer" for example become insignificant when compared to uncertainties in input data.

2.9.2 Terrestrial Model (O. Hoffman, R. Gardner)

A comparison of three models showed that the simplest one parameter model, gave results which were comparable to one of the more complicated models. A comparison of the results of the more complicated models showed quite a difference which was not readily explainable. Preliminary conclusions drawn from the comparison would suggest that the simple model is just as adequate as the other models.

2.10 Overall Comments

In general there were no major disagreements with the points made in the various discussion papers with the exception of the interpretation of the output distributions by Kaplan.

Table 2.1

WORKSHOP ATTENDEES

W. C. Harrison	National Uranium Tailings Program
R. John	National Uranium Tailings Program
R. Holmes	National Uranium Tailings Program
B. E. Halbert	SENES Consultants Limited
D. B. Chambers	SENES Consultants Limited
M. L. Murray	SENES Consultants Limited
L. M. Lowe	SENES Consultants Limited
V. J. Cassaday	SENES Consultants Limited
J. M. Scharer	University of Waterloo
P. M. Reilly	University of Waterloo
T. Duever	University of Waterloo
S. King	Golder Associates
R. V. Osborne	AECL, Chalk River Nuclear Laboratories
B. Goodwin	AECL, Whiteshell Nuclear Research Establishment
T. Andres	AECL, Whiteshell Nuclear Research Establishment
F. O. Hoffman	Oak Ridge, Tennessee
R. H. Gardner	Oak Ridge, Tennessee
B.G. Blaylock	Oak Ridge, Tennessee
S. Kaplan	Kaplan & Associates Inc.

3.0 SUPPLEMENTARY NOTES ON ENDPOINT FOR ASSESSMENT BY RICHARD V. OSBORNE, AECL, CRNL

The basic radiation protection principles and their application have been discussed in other papers in this study. The purpose of these notes is to indicate what calculable quantities are needed in order to apply radiation protection principles.

We need to estimate quantities that are related to:
- protecting individuals,
- choosing between options.

Applying the principles leads us to two types of criteria:
- mandatory criteria and
- preference criteria

3.1 Mandatory Criteria

These criteria are intended to ensure that individuals are protected sufficiently. The first criteria is concerned with the maximum risk to a most-highly-exposed individual at any time in the future. Here assurance is needed that the risk to such an individual is less than some defined limit. We also need to assure that the frequency of anyone intruding into tailings is sufficiently low that the overall individual risk is less than some limiting value.

A second criteria will be concerned with the average individual dose in a region or globally. Here assurance will be needed that there will be no appreciable widespread effect at any time in the future from the tailings.

Thirdly, there will be one or more criteria related to the maximum concentration of chemicals in public waters, in food or in air. Assurance will be needed that the concentrations will remain less than federal and provincial limits.

3.2 Preference Criteria

The intention in applying these criteria is to determine the most reasonable way of managing tailings. If we assume that some kind of multi-attribute analysis is to be used then we need to estimate quantities related to:

- radiation doses to populations
- radiation doses to individuals,
- costs and performance of engineering structures or technologies needed for managing the tailings

3.3 Choice of Parameters

It is unlikely that single values of parameters will be used with most of the models in an assessment. Although some parameters could have fixed values; for example, the size of the given tailings area or the grade of ore, others may have to be distributed, for example, annual rainfall and concentration factors for particular radionuclides. Where the choice of parameters is almost entirely a matter of judgement (i.e., a hypothetical pathway is being considered, or very little field data is available), it may be best to run different assessments with single values rather that use an arbitrary distribution within an analysis. Such an approach might be taken, for example, if quite distinct climatic conditions are to be modelled. A general guide might be that the least number of parameters that seems reasonable should be distributed.

In the particular case of the parameters describing the characteristics of a critical group, then, by definition, these parameters should not be distributed. If they were to be, there would be the question of what is meant by the critical group.

3.4 Individual Risk Estimates

The basic output needed is a histogram of the distribution of annual dose at successive times in the future to a member of a critical group. Such histograms might be estimated for different critical groups and for separate pathways and radionuclides but the intention should be to obtain a composite histogram that represents the frequency distribution of annual doses at times in the future to the member of the particular group that is critical for the combination of all radionuclides by all pathways. This output can then be manipulated for use with the various criteria that may be appropriate for individual risk limitation. Four such criteria can be identified (for additional details see the paper by R.V. Osborne and P.J. Barry in SENES (1985)):

1) The mean risk may be calculated as the expectation value for the risk to an individual and its variation with time. (Risk is taken here as the probability of a health effect per unit dose multiplied by the value of dose multiplied by the probability of the dose being received. See the referenced paper for a more detailed discussion.) The mean could be averaged over successive 50-year periods. This value could be compared with a risk limit or risk upper bound.

2) The probability of exceeding a given annual dose at various times in the future could be estimated. This is the downward cumulative probability distribution or the complementary cumulative probability distribution and it can be used in several ways. A criterion which is equivalent to No. 1 above would be to require that the area under a downward cumulative probability distribution (multiplied by

the probability of a health effect per unit dose) is less than a given risk limit.

3) A reference distribution could be defined and the criterion could be that the assessed downward cumulative probability distribution should, at all levels of individual doses, be less than the reference distribution. If the reference distribution has an area equal to the risk limit, then this criterion is at the least as limiting as that of 1) and 2) above.

4) Single values could be defined of probability and dose below which the estimated downward cumulative probability distribution must lie. This criterion does not limit the total individual risk so that if used it would be to complement one of the other criteria, the intention being to limit the probability of particularly high doses being received.

3.5 Comparison of Tailings Management Options

The intention here is to determine when radiation risks are going to be as-low-as-reasonably-achievable. The method recommended is one of multi-criteria or multi-factorial analysis. Quantities that need to be estimated can be listed under the three headings noted above:

Population Related

- difference in collective doses between options
- differences in distributions of collective dose in time
- differences in distributions of collective dose in space (i.e. local, regional, global)
- differences in distributions of collective dose in individual dose

- certainty of estimated values
- collective doses from low probability events.

Individual Related

- maximum risk to individuals
- duration of the maximum risk
- the time in the future that the maximum risk occurs
- the average dose over various regions
- the dose resulting from various intrusion scenarios

Engineering Related

- costs dependent on institutional control
- incremental-cost effectiveness (dollars per man Sievert)

This list could be extended to include any radiological factors that may have an influence on any choice of management option.

These factors or attributes can be used in comparisons between various options. Some may be judged far more important than others. Therefore, the "best" estimates of the various quantities together with the uncertainties associated with them will be needed, rather than "conservative" values. The relative importance accorded to the factors will determine what effort need be put into determining between the various options for those factors.

REFERENCES

SENES Consultants Limited, 1985. "Approaches to Risk Assessment for Canadian Uranium Mill Tailings." Research Report Prepared for the National Uranium Tailings Program, CANMET, EMR, Ottawa.

DISCUSSION PAPER ON

THE INTERPRETATION OF A PROBABILISTIC ASSESSMENT OF THE ENVIRONMENTAL EFFECTS OF URANIUM MILL TAILINGS

prepared by

S. Kaplan
Kaplan & Associates Inc.

in association with

L.M. Lowe and D.B. Chambers
SENES Consultants Limited

June 1985

Abstract

The interpretation of the output of a probabilistic assessment and its use by regulatory agencies are important concerns. The paper discusses the purpose of probabilistic modelling and how the interpretation of the model output is dependent on how the distributions assigned to the model inputs are established. The distinction between probability and frequency is discussed. Various possible regulatory uses and requirements for probabilistic models are presented. The accuracy of the Monte Carlo technique used in probabilistic modelling is assessed by means of Bayes' theorem. Numerical examples illustrating how to estimate statistical confidence in the model output are provided.

TABLE OF CONTENTS

Page

1.0 INTRODUCTION 1

2.0 STRUCTURE OF THE SENES MODEL - A SIMPLIFIED VERSION 2

3.0 INTERPRETATIONS OF THE MODEL 4
  3.1 Distinction Between Probability and Frequency 5
  3.2 Distinction between Probability Distributions and Frequency Distributions 6
  3.3 Uncertainty About Variability 7
  3.4 Three Possible Interpretations of the Model 9
  3.5 Accounting for Correlations Among Variables 11

4.0 REGULATORY USE OF THE MODEL 13
  4.1 Non-Robustness of the Tails 15
  4.2 A New Kind of Regulator? 16

5.0 ACCURACY OF THE MONTE CARLO TECHNIQUE 17
  5.1 Bayes' Theorem 17
  5.2 Bayesian Estimation of Shimmer, the Accuracy of Monte Carlo Results 19
  5.3 Evaluation of Bayes' Theorem in the Case of the Uniform Prior 21
  5.4 Numerical Examples of the Bayesian Approach 23

REFERENCES 26

List of Figures

		Follows Page
2.1	Simple Compartment Model	2
2.2	Flow of Calculations Showing "Backbone" and Feeder Lines	2
3.1	Variability Curve for Mean Annual Wind Speed	6
3.2	Probability Distribution for Wind Speed in 1976	6
3.3	Effect of New Information on Probability Curve	7
3.4	Probability Distribution for Wind Speed in "The Next" Year	7
3.5	Probability Distribution for Long-Term Average Wind Speed	9
4.1	Distribution for the Model Ouput Variables (X_N) Showing Regulatory Set Point ($\hat{X}_N$)	13
4.2	Two Distributions Having the Same Area (A) Above the Set Point ($\hat{X}_N$).	13
5.1	Output of an Infinite Monte Carlo Calculation	19
5.2	Posterior Probability	19

List of Figures (continued)

	Follows Page
5.3a Posterior Probability Density Functions of A for Various Sample Sizes m, Given k/m=0	24
5.3b Cumulative Probability Functions of A for Various Sample Sizes m, Given k/m=0	24
5.4a Posterior Probability Density Functions of A for Various Sample Sizes m, Given k/m=0.3	24
5.4b Cumulative Probability Functions of A for Various Sample Sizes m, Given k/m=0.3	24
5.5a Posterior Probability Density Functions of A for Various Sample Sizes m, Given k/m=0.1	25
5.5b Cumulative Probability Functions of A for Various Sample Sizes m, Given k/m=0.1	25
5.6a Posterior Probability Density Functions of A for Various Sample Sizes m, Given k/m=0.05	25
5.6b Cumulative Probability Functions of A for Various Sample Sizes m, Given k/m=0.05	25

1.0 INTRODUCTION

The present paper discusses a series of topics related to the probabilistic assessment of the environmental effects of Canadian uranium mill tailings. Most important among these topics are some conceptual issues related to the interpretation of the probabilistic technique and its possible use by regulatory agencies. In addition, some points connected with the formulation and computational procedures used in the Phase I SENES (1984) study are discussed.

The basic purpose of the SENES model is to aid decision making regarding the long-term management of uranium tailings piles. The model provides an explicit and quantitative basis for such decision making, and is, in fact, an excellent example of the usefulness of a mathematical modelling approach to difficult public health and environmental policy issues.

Critics will object that the model, like all mathematical models, is only an approximation to physical reality. One response to this objection is to include, as SENES does, a treatment of uncertainty within the model itself. Thus, instead of delivering as output a single number of unknown accuracy, the model delivers to the decision maker an entire probability distribution which expresses our whole state of knowledge about the number.

At first glance, this fuller output may seem to complicate the decision makers' problem. Properly understood however, it greatly simplifies decision making by explicitly exposing the whole truth: what is known and what is not known. Indeed, this probabilistic aspect of the model is no more than traditional good scientific practice applied to the decision making situation, an area where it is much needed.

To those critics who object more fundamentally, who are uncomfortable with a mathematical approach per se, we answer simply by saying "What's the alternative?" Without a quantitative framework, the debates about health and environment degenerate into shouting matches, demagoguery and prolonged litigations which bring us nowhere. With a rational and quantitative framework, we have at least the possibility of achieving understanding, communication and constructive action.

The purpose of the present paper is to further this action by improving the understanding and usefullness of probabilistic assessment techniques as they may be applied to uranium mill tailings. For this purpose, we shall begin by describing a very simplified version of the Phase I model structure and by presenting some elementary, but very important, ideas from probability theory. In so doing we may cover some of the same ground treated in the Phase 1 report (SENES 1984). Hopefully this repetition, from a slightly different angle, will be as useful to the reader as it has been to the writer. In addition, the possible regulatory use of the model is discussed as are questions related to the accuracy of Monte Carlo calculations.

2.0 STRUCTURE OF THE SENES MODEL - A SIMPLIFIED VERSION

The model used by SENES in their Phase I studies to trace the flow of contaminants from the tailings pile, is of a type known as a "compartment model" because it divides the world into a finite set of "compartments" representing such entities as "fish", "garden vegetables", "top metre of the tailings pile", etc. In the simplest case, these compartments are arranged in a chainlike fashion, as shown in Figure 2.1.

The material in question, in this instance radium-226, is fed

FIGURE 2·1

SIMPLE COMPARTMENT MODEL

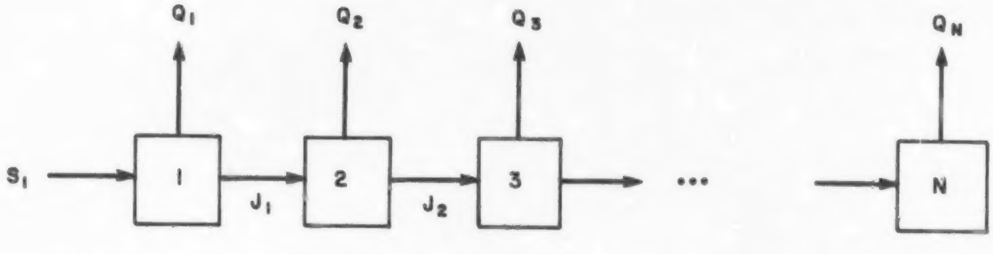
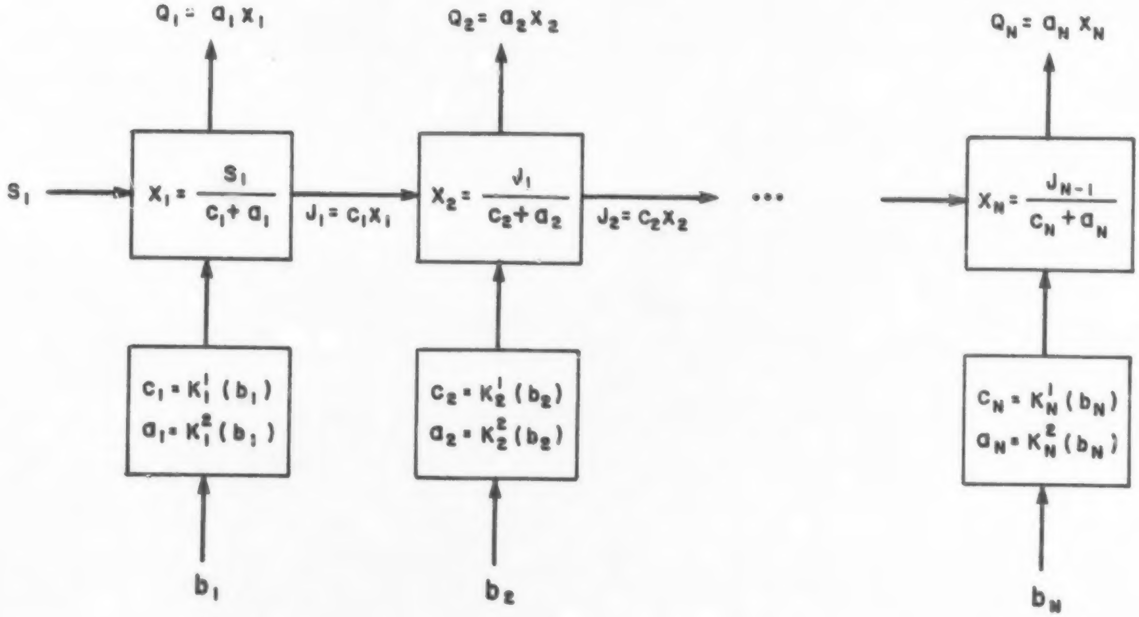


FIGURE 2·2

FLOW OF CALCULATIONS

SHOWING "BACKBONE" AND FEEDER LINES



into compartment 1, at the left, with flow rate S_1 (denoting the source). The flow rates from compartment 1 to compartment 2, compartment 2 to 3, etc., are designated J_1, J_2, etc. Each compartment has a "content" of material, which is designated by x_1, x_2 .... x_N and a "side flow" or "absorption" which is denoted by Q_1, Q_2 ... Q_N. The side flows exit the model and can be taken to represent, for example, radioactive decay.

We seek to know the flows and contents of each compartment, but most particularly the Nth compartment, which represents the final target, e.g. man. For this purpose, it is necessary to state the relationship connecting flows and contents. The SENES model uses a linear relationship and is therefore a "linear" compartment model. That is, it assumes that the flows from a compartment are proportional to the content of the compartment. Thus for any compartment "i", it can be stated that:

$$J_i = c_i x_i \qquad (2\text{-}1)$$
$$Q_i = a_i x_i$$

where we shall call c_i the "communication coefficient" and a_i the "absorption coefficient" of compartment i. Assuming steady state, it is possible to immediately calculate from equations (2-1) the contents of all compartments by applying conservation of mass from left to right. Thus:

$$\begin{aligned} x_1 &= \frac{S_1}{c_1+a_1}, & J_1 &= c_1 x_1 \qquad (2\text{-}2) \\ x_2 &= \frac{J_1}{c_2+a_2}, & J_2 &= c_2 x_2 \\ &\vdots & &\vdots \\ x_N &= \frac{J_{N-1}}{c_N+a_N} \end{aligned}$$

This gives us the solution for the flows and contents in terms of the compartment coefficients c_i, a_i. In the Phase I model, these coefficients are themselves calculated from more basic parameters using detailed models of the individual compartments. To represent these calculations we write:

$$c_i = K_i^1 (b_i)$$

$$a_i = K_i^2 (b_i) \qquad (2\text{-}3)$$

where the b_i parameters represent the basic parameters of the ith compartment model and the K_i may be called the "compartment functions".

If these individual compartment models are included in the diagram, Figure 2.2 is obtained, in which the overall flow of the calculations can be viewed as a backbone or assembly line fed from below by the tributary lines of the individual compartment models.

3.0 INTERPRETATION OF THE MODEL

Figure 2.2 shows the structure of the model. In computing with this model, it is recognised that the input parameters, b_i's, may not be constant in time, and even if they are constant, their actual numerical values are uncertain. Therefore, the SENES model inputs distributions for these parameters and, through a Monte Carlo procedure, calculates distributions for the output quantities, particularly x_N. The question therefore arises as to how these output distributions are to be interpreted for purposes of decision making.

The interpretation of the output distributions depends upon exactly how the input distributions are established. To discuss this point further, the distinction between the concepts of variability and uncertainty must be clarified. For this purpose we shall begin by defining a distinction between the words "frequency" and "probability" as we shall use them in this paper.

3.1 The Distinction Between Probability and Frequency

People have been arguing about the meaning of probability for at least 200 years, since the time of Laplace and Bayes. The major polarization of the argument is between the "objectivist" or "frequentist" schools who view probability as something external, the result of repetitive experiments, and the "subjectivists" who view probability as an expression of an internal state - a state of knowledge or a state of confidence.

This paper adopts the point of view that both schools are right; they are just talking about two different ideas. Unfortunately, they both use the same word - which seems to be the source of most of the confusion. What needs to be done therefore is to give each idea the dignity of its own name. In the present paper, we shall do this by calling one idea frequency and the other probability.

What the objectivists are talking about we call "frequency". What the subjectivists are talking about we call "probability". Thus, probability, as we use it, is a numerical measure of a state of knowledge, a degree of belief, a level of confidence. Frequency, on the other hand, refers to the outcome of an experiment of some kind involving repeated trials. Thus frequency is a "hard" measurable number. This is so even if the experiment is only a thought experiment or an experiment to be done in the future. At least in concept then, a frequency is a well-defined, objective, measurable number.

Probability, on the other hand, is a notion of a different kind. Defined as a number used to communicate a state of mind, it is thus inherently subjective and changeable as new information arrives. To make this notion useful, it is obviously important to clearly define the correlation between the numbers and the state of mind. We shall not stop to do this here. Suffice it to say that it can be done with complete satisfaction. The scale of probability, i.e. the scale of confidence, can be defined as rigorously as the scale of length or temperature or anything else, and from this definition, the usual axioms and rules of probability theory fall out in a natural way.

3.2 The Distinction Between Probability Distributions and Frequency Distributions

Corresponding to the above distinction between probability and frequency, it is also necessary to make a strong distinction between probability and frequency "distributions". For this purpose consider the input parameter: "mean annual wind speed above the tailings pile". This parameter varies from year to year. Thus if we had data over many years, a frequency distribution curve showing this variability could be drawn, as shown on Figure 3.1 for example. The area under this curve over any interval u, v shows the fraction of years that the measured wind speed fell in this interval.

Now suppose we next asked "What was the wind speed in 1976?" Unless we had the data in front of us, we would be uncertain about the answer. We would therefore express our uncertainty in the form of a probability distribution such as that shown in Figure 3.2. The area under this curve over the interval u, v is our degree of confidence (as measured by what odds we would take) that the actual mean speed in 1976 was between u and v. (We would have the same degree of confidence about what the wind speed was

FIGURE 3·1

VARIABILITY CURVE FOR MEAN ANNUAL WIND SPEED

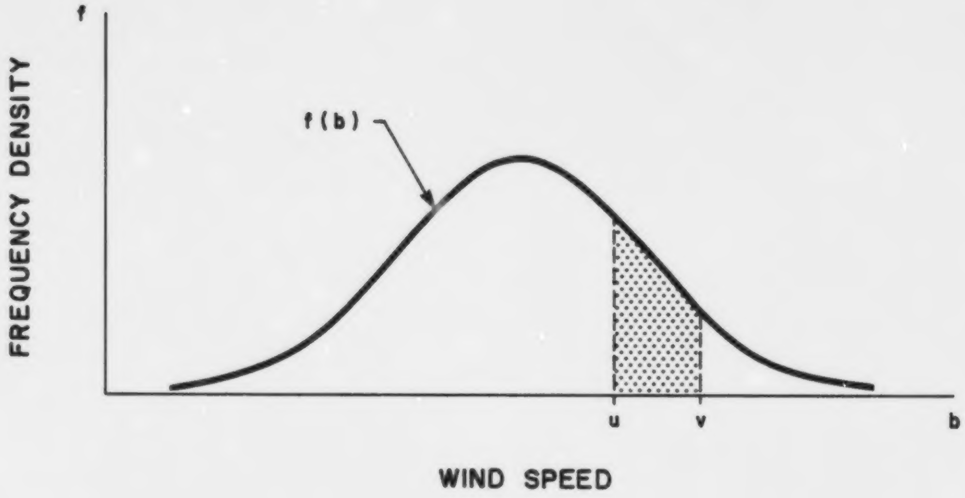
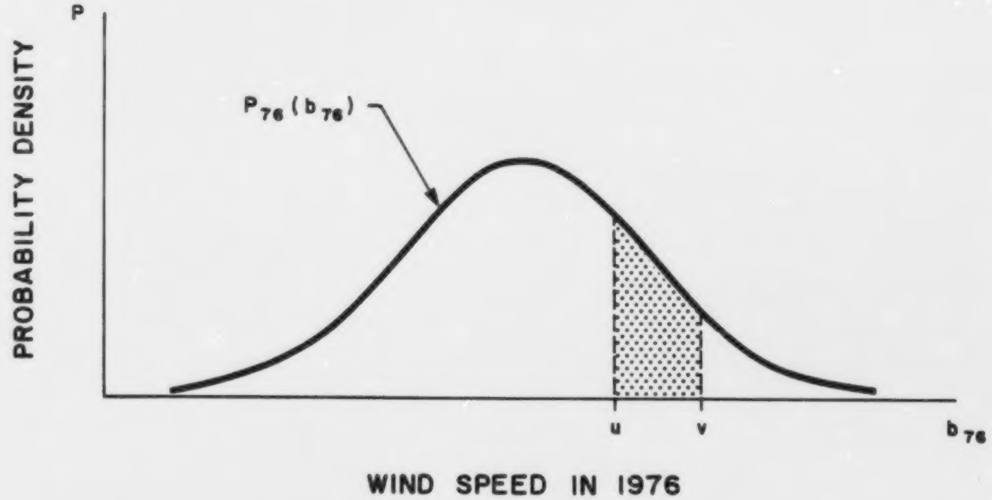


FIGURE 3·2

PROBABILITY DISTRIBUTION FOR WIND SPEED IN 1976



in any year.)

Note that the curve in Figure 3.2 would have precisely the same shape as that in Figure 3.1. It would be numerically identical but conceptually it is a very different animal. Figure 3.1 is a frequency distribution against a true random variable, b, which fluctuates from year to year. Figure 3.2 is a probability distribution against a fixed (but uncertain) variable, b_{76}. Figure 3.1 shows the variability of the fluctuating variable b and thus portrays a phenomenon that exists in the real world. Figure 3.2 portrays our uncertainty about the fixed quantity b_{76} and thus portrays something that exists only in our heads.

Our state of certainty of course changes when our information changes. Thus, for example, suppose we found out somehow that the year 1976 was one of the windiest on record. Our probabilistic estimate of b_{76} would now shift to something like that shown in Figure 3.3. Figure 3.1 of course would not change at all.

3.3 Uncertainty About Variability

Now let us take this discussion one step further. In Figure 3.1, it was assumed that there were many, many years of data from which to draw the variability curve. Suppose now that only limited data, say 10 or 20 years' worth are available. From this limited sample, an approximate variability curve, $\tilde{f}(b)$, could be drawn, but it would be important to recognise that this would not be the same as the "true" curve, f(b), that would be drawn after hundreds or thousands of years of data.

With the limited sample, we are thus in a position of being uncertain about the variability. In other words, the function f itself is uncertain. If desired, this uncertainty could actually

FIGURE 3·3

EFFECT OF NEW INFORMATION
ON PROBABILITY CURVE

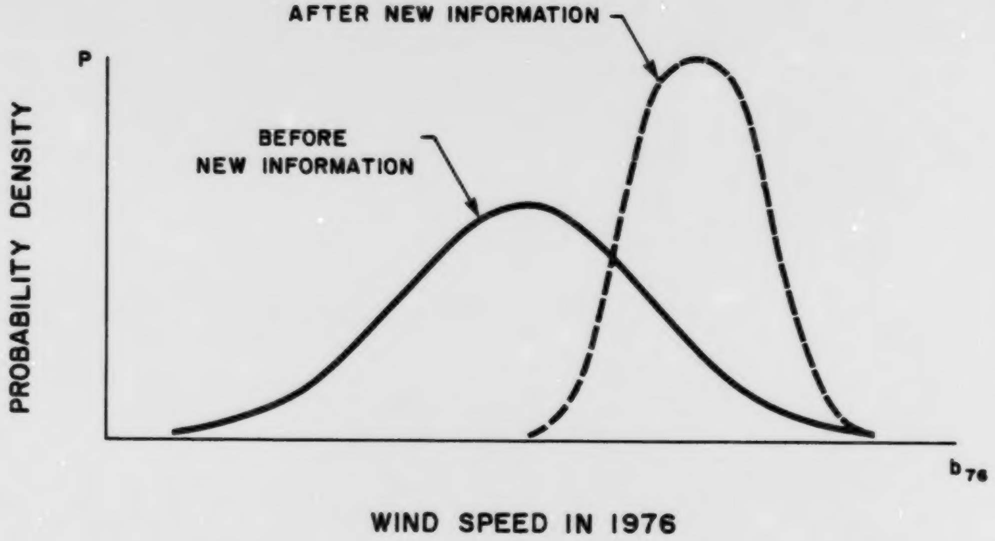
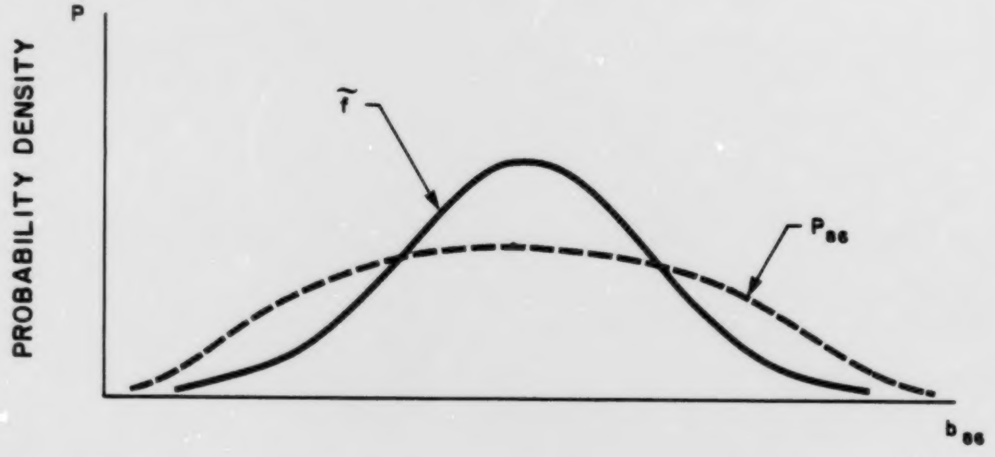


FIGURE 3·4

PROBABILITY DISTRIBUTION
FOR WIND SPEED IN "THE NEXT" YEAR



be quantified by erecting a probability distribution, $p_f(f)$, over the space of possible functions, f, (see Kaplan (1983) for examples of this).

Suppose then that we were in this position of uncertainty about variability (which is where we usually are) and suppose we now ask "What will be the mean wind speed for next year, i.e. 1986?" To answer this, a probability distribution must be provided which includes both the uncertainty about b_{86} stemming from the year to year variability and also our uncertainty about what that variability is. The distribution which accomplishes this purpose is:

$$p_{86}(b_{86}) = \int_f p_f(f)\, f(b_{86})\, df \qquad (3\text{-}1)$$

In this equation, $f(b_{86})$ is playing the role of a conditional probability:

$$f(b_{86}) = p(b_{86}|\, f) \qquad (3\text{-}2)$$

That is, given that the true frequency distribution is f, the probability that the wind speed in 1986 will have the particular value b_{86} is numerically equal to the value of the function f at b_{86}.

Equation (3-1) thus clearly shows that we have two sources of uncertainty about b_{86}. One source is the year to year variability, expressed by the function f(b). The second source is our incomplete knowledge about the variability, expressed by the function $p_f(f)$. Because of these two sources of uncertainty the curve for p_{86}, plotted from equation (3-1), must be broader than the experimental curve $\tilde{f}$, as illustrated in Figure 3.4.

One further point should be noted. Suppose now we were interested only in the average wind speed, b, over the lifetime of the tailings pile, say 1000 years. If we knew the variability curve we would calculate this average as:

$$\bar{b} = \int_0^\infty b f(b)\,db \qquad (3\text{-}3)$$

and there would be essentially no uncertainty in this average. On the other hand, if we have uncertainty about f this would lead directly to uncertainty about $\bar{b}$. Symbolically:

$$p_f(f) \rightarrow p_{\bar{b}}(\bar{b}) \qquad (3\text{-}4)$$

An example of this symbolic relationship can be seen most easily using the concept of Discrete Probability Distributions (DPD's) (Kaplan 1981). From this point of view ,the DPD for f is:

$$f = \{\langle P_i,\ f_i \rangle\}$$

which translates directly into a DPD for b as follows:

$$b = \{\langle p_i,\ \bar{b}_i \rangle\} \text{ where } \bar{b}_i = \int_0^\infty b_i f_i(b)\,db$$

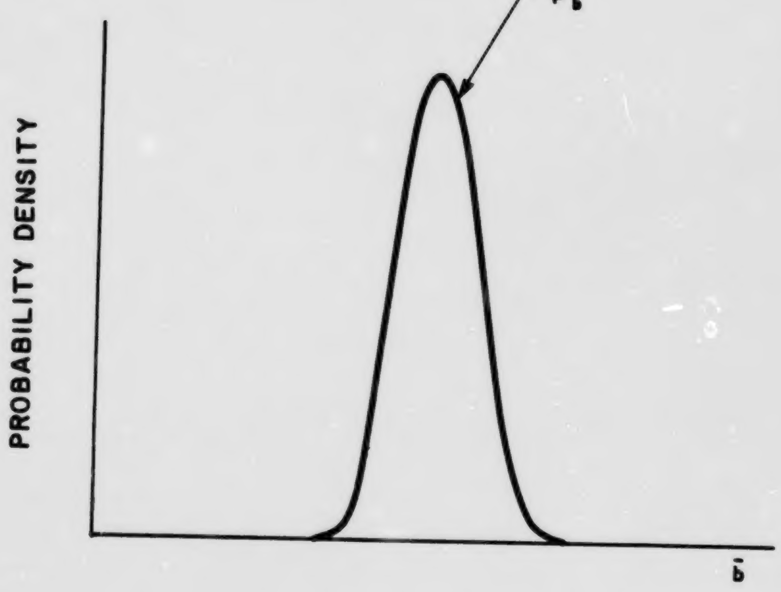
The distribution $p_{\bar{b}}$ would be relatively narrow as shown in Figure 3.5, since it reflects only the uncertainty in the variability and not the variability itself. Saying this another way, Figure 3.2 represents uncertainty resulting from variability, Figure 3.5 represents the uncertainty about the variability, and curve p_{86} in Figure 3.4, represents the combination of those two uncertainties.

3.4 Three Possible Interpretations of the Model

We return now to the question of interpreting the distribution of

FIGURE 3.5

PROBABILITY DISTRIBUTION
FOR LONG TERM AVERAGE WIND SPEED



the output, x_N, of the model. Three possible different interpretations can be distinguished.

Interpretation 1 - If all the input distributions for S_1, and the b_i parameters, are meant as pure variability curves (like Figure 3.1), then the output distribution for x_N is also a variability curve. It has the interpretation therefore of showing in what fraction of future years the value x_N will lie in any given interval.

Interpretation 2 - Suppose we were primarily interested in characterizing the average performance of the tailings pile over a long period, say the first 1000 years of its life. We would then define the parameters in the model, S_1 and b_i, as average values over this period. Our input distributions should then be pure probability distributions, like Figure 3.5, expressing our uncertainty about these average parameter values. In this case the output quantity x_N would then also be interpreted as the average value of compartment N content over the first 1000 years. The output distribution for x_N would be interpreted as a pure probability distribution expressing our state of confidence about that average value.

Interpretation 3 - If all input distributions are meant, in the sense of Figure 3.4, as probability distributions for "the next" year (or equivalently for a randomly chosen year) then the output distribution must similarly be interpreted as showing our state of confidence that the value of x_N in the next year will fall within any given interval.

Until now, the Phase I model appears to have been used with interpretation 3. The present author, out of long habit, prefers to think in terms of interpretation 2. The information to allow us to use interpretation 1 is not available. If some parameters are input under one interpretation, and others under a different interpretation, then the output cannot be interpreted in a pure form of 1,2 or 3, and indeed may not be interpretable. Thus, for example, suppose we are primarily interested in the long term behaviour of the pile, specifically say the average behaviour over the first 1000 years. Then the distribution we input for the "mean annual wind speed " should not reflect, as Figure 3.4 does, the year to year variability of this parameter. That distribution would be too broad. The one we input should be narrow, like Figure 3.5, reflecting only our uncertainty in the 1000 year average.

As a practical matter, if in the important input parameters the uncertainty dominates over the variability then interpretation 3 collapses effectively to interpretation 2.

3.5 Accounting for Correlations Among Variables

In all probabilistic calculations attention must be paid to the question of correlated variables. First, as to the meaning of this term, suppose we have two variables, say b_1 and b_2, and we have set forth for each a probability distribution giving our state of knowledge about its possible values. Now suppose somehow we find out the precise value of b_1. If this information causes us to change our probability curve for b_2 then we say the two variables are correlated, or have correlated uncertainties.

They are also then said to be "dependent". If learning the value of one variable has no effect on our knowledge of the other then the variables are "uncorrelated" or "independent".

It is important to keep track of correlations because, for example, if two positively correlated variables are added as if they were uncorrelated, then the resulting distribution is much too narrow, indicating a greater degree of certainty than we really have in the value of the sum.

The best way to keep track of correlations, in the writer's experience, is to explicitly identify the root variables that are the source of the correlation. Thus, for example, in a construction project, suppose x is the cost of labour and y the cost of materials. The total cost z is then:

$$z = x + y \qquad (3\text{-}3)$$

Now suppose we have uncertainty in both x and y, and suppose these uncertainties are partially correlated through the inflation rate, r. That is to say, suppose both x and y depend upon r but also upon other variables u and v which are not correlated. Thus

$$\begin{aligned} x &= f(u,r) \qquad (3\text{-}4) \\ y &= g(v,r) \end{aligned}$$

Then it is best to write equation (3-3) in the form:

$$z = f(u,r) + g(v,r) \qquad (3\text{-}5)$$

In this equation, the root variables u,v and r are all independent. Therefore, z can be calculated from this equation using the usual Monte Carlo or Discrete Probability methods for combining independent variables.

4.0 REGULATORY USE OF THE MODEL

The traditional regulatory approach, in Canada and elsewhere, is to set a limiting value and require that the variable being regulated stay below that limit. This way of thinking has perhaps served well for automobile speed limits, concentrations of food additives, etc. However, this approach is no longer adequate for all applications. Attempts to apply it to regulation of modern technology and waste disposal run into severe conceptual and philosophical difficulties because this mode of thinking neglects the fundamental decision theoretic character of regulatory actions (Kaplan, 1982). In addition to fundamental philosophical difficulties, further serious mechanical and conceptual difficulties arise when this approach is applied to situations involving variability and uncertainty. In this section, we shall point out some of these latter difficulties.

Suppose, for example, it is desired to place a limiting value, $\hat{x}_N$, on the content x_N of the last compartment. Suppose further that the output of the model is a distribution which, as illustrated in Figure 4.1, has the shaded area, A, above $\hat{x}_N$. Then, under the previous three interpretations, we would conclude:

Interpretation 1 -	In the fraction, A, of future years, the content of compartment N will exceed the regulatory limit.
Interpretation 2 -	The probability (i.e. our degree of confidence) that the average content of compartment N during the first 1000 years will exceed $\hat{x}_N$ is equal to A.

FIGURE 4·1

DISTRIBUTION FOR THE MODEL OUTPUT VARIABLE (x_N) SHOWING REGULATORY SET POINT ($\hat{x}_N$)

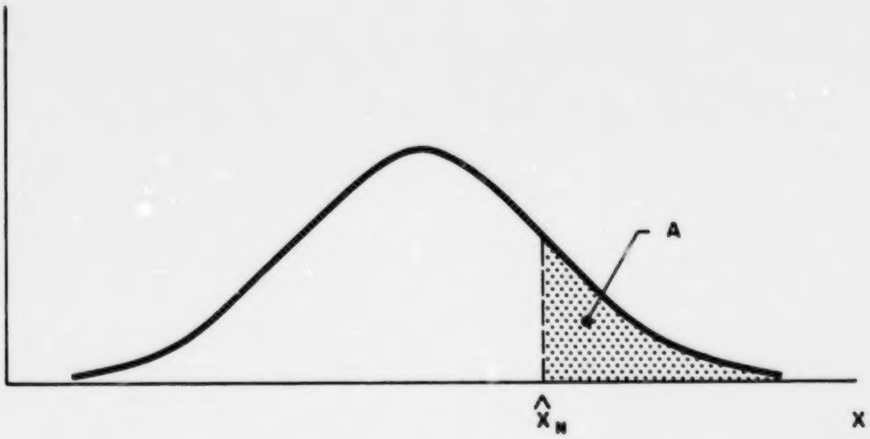
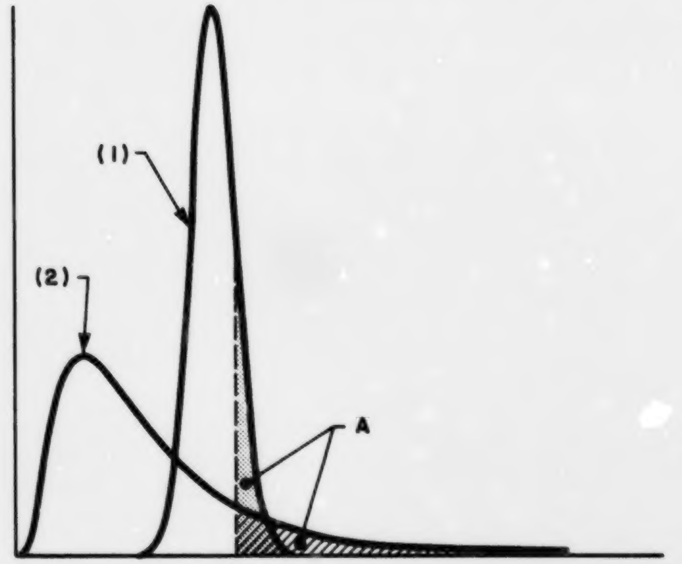


FIGURE 4·2

TWO DISTRIBUTIONS HAVING THE SAME AREA (A) ABOVE THE SET POINT ($\hat{x}_N$)



Interpretation 3 - The probability that x_N will exceed $\hat{x}_n$ next year is A.

Under interpretation 1, we would be moved then to ask ourselves whether the fraction A was acceptable and thence to expand the regulation to a statement like: The fraction of years in which x_N exceeds $\hat{x}_N$ shall be no greater than $\hat{A}$.

Similarly, under interpretations 2 and 3, applying the regulatory approach would lead us to set a limit, $\hat{A}$, on the "admissible" value of the shaded area, A.

A Conceptual Difficulty

The above regulatory approach to limiting the value of A is conceptually unsatisfactory for, among other reasons, it would label the two curves in Figure 4.2, both having the same area A beyond $\hat{x}_N$, as equally desirable (or undesirable). But actually these two curves have very different implications for decision making. For example, under interpretation 2, curve (1) of Figure 4.2 would tell us that the tailings pile is unacceptable as designed and that something physical should be done to change it. Curve (2) however, would tell us that the pile most probably is very well designed and will perform admirably. However, we are not quite sure of that. Curve (2) would tell us then, not to change the pile physically, but to study it further, to do more research so as to improve our level of confidence.

One attempt to cure the above difficulty is to prescribe a series of limiting values $\hat{x}_N{}^1$, $\hat{x}_N{}^2$, $\hat{x}_N{}^3$ ..., each with its own exceedance probability $\hat{A}^1$, $\hat{A}^2$, $\hat{A}^3$, etc. But this leads to an excessively cumbersome regulation, and also runs up against other "mechanical" problems as discussed below.

4.1 Non-Robustness of the Tails

The natural tendency of a regulator would be to set values of A in the range of 5% or less. This places the whole burden of the regulatory pass/fail decision on the detailed shape of the upper tail of the output distribution. But this tail is a very non-robust part of the calculation. "Non-robust" means that it is changeable, unstable, not meaningful. Essentially the shape of the tail is an artifact of the calculation.

The reasons for this non-robustness are twofold. First, the extreme tail of the output distribution is dependent on the tails of the input distributions. But the tails of the input probability distributions are poorly "known", that is to say, they are not well dictated either by evidence or opinion.

Second, even if the input tails were meaningful and well defined, there are computational difficulties in translating these tails to the tail of the output distributions. Thus the tail of the output distributions also contains a purely computational error or computational uncertainty, particularly if a conventional Monte Carlo procedure is used in the calculations. (See section 5.4 for numerical examples of this.)

In short, when doing probabilistic work one should keep in mind that it is the general, overall shape of the probability curve which contains the information about our state of knowledge. Fine details of the shape, especially in the tails, are usually not significant. They usually do not carry information and usually should not be taken seriously.

4.2 A New Kind of Regulator?

Well, if setting limits $\hat{x}_n$ and $\hat{A}$ are unsatisfactory, then what is a poor regulator to do? The answer is that he must look at the whole curve. The whole curve tells the story, not any parameter of the curve $p(x_N)$ such as mean or median or mode or variance or percentile - not any combination of such parameters either. The whole curve tells the whole story. Why should one look at less than the whole story? Why should one throw information away when making a decision?

The decision maker (i.e. the regulator) must learn the language of curves. He must study the whole curve and read its messages. He must understand why the curve has the particular shape it has and where this shape comes from. For this purpose he must study the curves for the intermediate quantities x_i, J_i, a_i, c_i, Q_i and the input quantities S_1 and b_i. The model should facilitate the decision maker's task by presenting these intermediate curves, and comparisons of these curves, in convenient ways. But the decision maker also must be willing to dig. Only in that way can he gain deep understanding. And only then can he propose changes to the design or management of the pile, calculate new curves, and ultimately make the judgements about what changes if any are warranted. Decision making on a subject like this should not be simply a matter of meeting goals or criteria. It should be a matter of responsible judgement with proper balancing of all safety, environmental and economic issues.

The simplistic regulatory approach of the past must now evolve into a "holistic" approach, one that recognises that making regulations is tantamount to making decisions, and that these decisions can have far reaching effects on our society. The regulator must therefore recognise himself as decision maker, acting as agent for the society. As such then, he must be guided by the conceptual structure of Decision Theory, which structure requires that in a decision analysis we examine fully all the possible alternatives and all the aspects of each alternative, i.e. the costs, benefits, risks, uncertainties and preferences.

5.0 **ACCURACY OF THE MONTE CARLO TECHNIQUE**

As already noted, the model uses distributions for the input variables S_1 and b_i, and calculates distributions for the output variables using the Monte Carlo technique. That is, it randomly selects a set of input values, calculates the output, (say x_N), and repeats the process a designated number (say m) times. Since m will be a finite number (typically in the order of 10^2 or 10^3), there will be a sampling error from the Monte Carlo technique itself. This sampling error has been designated "shimmer" (Andres 1984a), and introduces an additional computational uncertainty to the uncertainty and variability already built into the model. The question therefore arises as to how one might estimate the magnitude of this "shimmer" error. This question has been addressed by Andres (1984b) using the customary statistical concept of confidence interval estimation. In the following sections of this paper, the question is addressed in a related but different way using Bayes' theorem and leading to what could be called Bayesian confidence intervals. First however, let us consider what the "Bayesian" approach is. In section 5.4, we shall give numerical examples.

5.1 Bayes' Theorem

The essence of the "Bayesian" approach is the interpretation of "probability" set forth in Section 3.1. In this interpretation, probability is a numerical scale that we establish to express and communicate our level of confidence with respect to any given proposition. In this interpretation therefore, probability does not measure something in the "real" or "outside" world; it measures rather an "internal" state, a state of confidence. By its definition therefore probability is inherently "subjective".

On the other hand, our level of confidence in any proposition is based upon, and logically related to, the evidence we have relevant to that proposition. But evidence is an objective thing; we can all agree on what the evidence is. Thus if we all have the same total body of evidence then, since we all ultimately process evidence in the same logical way, we should all arrive at the same level of confidence, i.e. the same probability. In this way we see therefore that probability is also "objective". It is really a numerical expression of, or encoding of the evidence and information we have.

It remains to explain the logical process by which we encode or evaluate our evidence in the form of probability. For this purpose the key idea is a celebrated, and much misunderstood and maligned theorem published posthumously by the Reverend Thomas Bayes in the 18th century.

Bayes' Theorem is the fundamental law of logical inference, the fundamental rule for the evaluation of evidence. To derive it requires only two steps. We begin with a fundamental property of probability:

$$p(A \text{ and } B) = p(A)\ p(B|A) \qquad (5\text{-}1)$$

i.e. the probability of propositions A and B both being true is equal to the probability that A is true times the conditional probability that B is true, given that A is true.

This property follows from the fundamental way in which we define probability as a scale of levels of confidence. In the same way then that we wrote equation (5-1), we may write.

$$p(A \text{ and } B) = p(B)\ p(A|B) \qquad (5\text{-}2)$$

Thus, equating the right sides of equations (5-1) and (5-2) and dividing by p(B) we obtain.

$$p(A|B) = p(A) \left[\frac{p(B|A)}{p(B)}\right] \qquad (5\text{-}3)$$

This is Bayes' Theorem. It says that $p(A|B)$, our level of confidence about A after we learn B, is equal to our level of confidence, p(A), before learning about B, times a correction factor, the term in brackets. The term in brackets therefore "updates" our level of confidence in A, based on the new evidence B. (When event B can take place jointly with one of several mutually exclusive events A_i, p(B) can be written as $p(B) = \sum_i p(B|A_i)p(A_i)$ in Bayes theorem.)

Thus we have derived Bayes' Theorem. Now let us see how to apply it to the question of assessing shimmer; that is, to assessing our state of confidence based on the results of a finite Monte Carlo calculation.

5.2 Bayesian Estimation of Shimmer, the Accuracy of Monte Carlo Results

Suppose we had conceptually run a Monte Carlo calculation with an infinite number of trials and had obtained for x_N the distribution, $p_N(x_N)$, and its integral, or cumulative form, $P_N(x_N)$, shown in Figure 5.1.

FIGURE 5·1

OUTPUT OF AN INFINITE MONTE CARLO CALCULATION

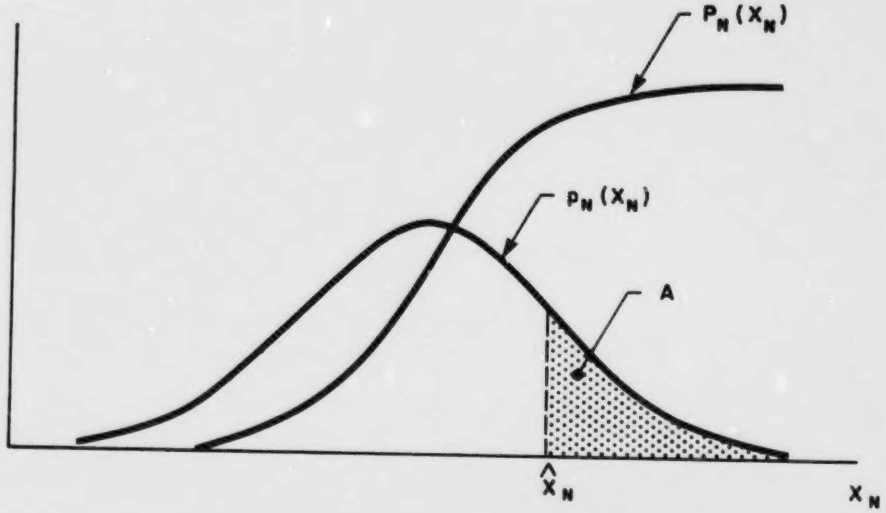
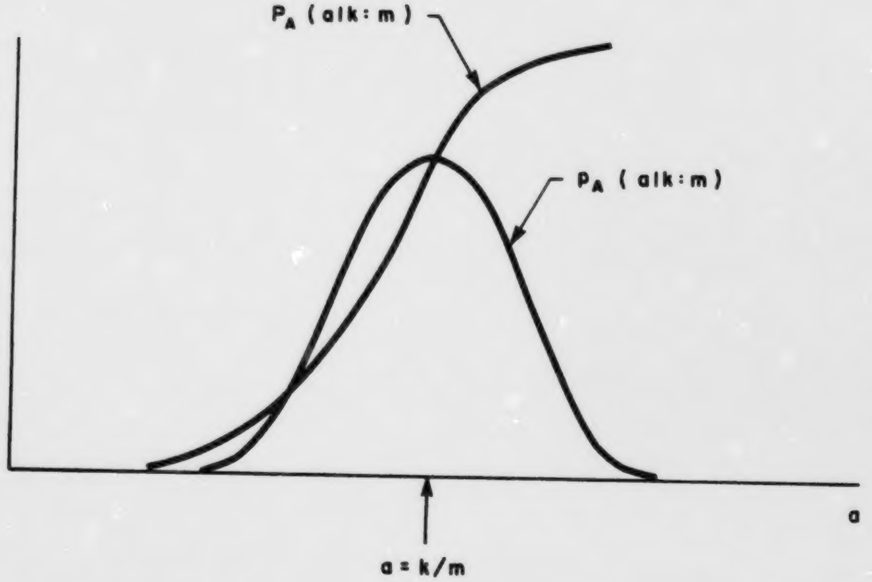


FIGURE 5·2

POSTERIOR PROBABILITY



The relation between P_N and p_N is of course:

$$P_N(x_N) = \int_0^{x_N} p_N(x)\, dx \tag{5-4}$$

Again, let A be the fraction of area exceeding the set point x_N.

$$A = \int_{x_N}^{\infty} p_N(x)dx = 1-P_N(\hat{x}_N) \tag{5-5}$$

Suppose we next did a finite Monte Carlo run with m trials, out of which k trials had results $x_N > \hat{x}_N$. The question now is: "From the evidence of this finite run, what inferences can be drawn about the "true" value of A, that is about the value of equation (5-5), that would be obtained from an infinite experiment?"

This is the way a Bayesian would pose the question of Monte Carlo accuracy. To answer this question, consider Bayes' Theorem in the form:

$$p_A(a|k:m) = \frac{p_A(a) \quad p_A(k:m|a)}{\int_0^1 p_A(a)\ p_A\ (k:m|a)da} \tag{5-6}$$

The term on the left here is the probability density for A conditional on the evidence of k occurrences in m trials. In other words,

$$p_A(a|k:m)da \tag{5-7}$$

is our level of confidence, given the information k out of m, that the true value of A lies in the interval da about a. This is the quantity we are looking for. It is called the "posterior" probability, i.e. the probability "after" learning the information k out of m.

Similarly, $p_A(a)$ is the probability density for A before learning the results of the Monte Carlo run. It is the "prior" probability.

The other term in the numerator, $p_A(k{:}m|a)$, is called the "likelihood function". It is the probability, if the true value of A were a, that we would have observed in the Monte Carlo run what we did observe, namely k occurrences out of m trials. It expresses therefore how likely is the observed result given the premise that A=a.

Bayes' Theorem, equation (5-6), thus tells us that our level of confidence that A=a, after learning the result k:m, is equal to our level of confidence prior to learning this result, multiplied by the likelihood of this result given A=a. The product is then normalized by integrating over all possible values of a.

Equation (5-6), plotted as a function of a, would result in a curve peaking in the neighbourhood of a = k/m, with the width of the curve decreasing as m increases, as illustrated in Figures 5.3 to 5.6. From these figures, we can obtain whatever confidence intervals we want.

5.3 Evaluation of Bayes' Theorem in the Case of a Uniform Prior

There is a large amount of literature and often vigorous controversy on the subject of specifying the prior distribution. It would be a distraction to discuss this subject here. We shall limit ourselves therefore to stating that in our opinion this controversy stems from a misunderstanding of probability. When a probability distribution is properly understood as an expression of our state of knowledge, then specifying the prior becomes an easy and most natural thing to do.

For the purpose of the present section, we shall specify a uniform prior, mainly for the sake of making our computations easily reproducible by the reader. In specifying a uniform prior, we are saying in effect that prior to seeing the Monte Carlo results we considered all values of A (between 0 and 1 of course) to be equally likely. To evaluate equation (5-6) for this case, we note, as does Andres (1984b), that the likelihood is a binomial function:

$$p_A(k:m|a) = \frac{m!}{k!(m-k)!} a^k (1-a)^{m-k} \qquad (5\text{-}8)$$

In the case where we have no knowledge of the value of A, prior to the Monte Carlo run, we set $p_A(a)=1$, and note that the integral:

$$\int_0^1 p_A(k:m|a)\,da = \frac{m!}{k!(m-k)!} \int_0^1 a^k (1-a)^{m-k} da$$

$$= \frac{m!}{k!(m-k)!} B(k+1, m-k+1) \qquad (5\text{-}9)$$

where B(u,v) is a beta function (Rinehart and Company 1965):

$$B(u,v) = \frac{(u-1)!(v-1)!}{(u+v-1)!} = \frac{\Gamma(u)\,\Gamma(v)}{\Gamma(u+v)} \qquad (5\text{-}10)$$

Thus from equations (5-9) and (5-10):

$$\int_0^1 p_A(k:m|a)da = \frac{m!}{(m+1)!} = \frac{1}{m+1} \qquad (5\text{-}11)$$

which used in equation (5-6) gives our desired result:

$$p_A(a|k:m) = \frac{(m+1)!}{k!(m-k)!} a^k(1-a)^{m-k} \qquad (5\text{-}12)$$

It is of interest additionally to note that the cumulative posterior:

$$\begin{aligned} P_A(a|k:m) &= \int_0^a p_A(a|k:m)da \qquad (5\text{-}13) \\ &= \frac{(m+1)!}{k!(m-k)!} \int_0^a a^k(1-a)^{m-k}da \end{aligned}$$

is the incomplete beta function which has been tabulated by Pearson (1948).

It is important to note that equations (5-12) and (5-13) are independent of the model used to generate the output for the Monte Carlo runs.

5.4 Numerical Examples of the Bayesian Approach

In the previous sections we have emphasized the role of Bayes' Theorem as a tool (actually "the" tool) for evaluating evidence, specifically the evidence from a finite Monte Carlo calculation. In this section, we shall give some numerical examples. It is particularly illuminating to show numerically how our state of knowledge about A changes as the evidence accumulates step by step.

For this purpose consider Figure 5.3a. Curve (0) shows our knowledge about A prior to running any Monte Carlo trials at all. The flatness of this curve expresses the fact that, prior to the runs, we consider all values of A to be equally likely. Now suppose we run one Monte Carlo trial, in which it turns out that x_N is below $\hat{x}_N$. Our evidence then is, $k = 0$, $m = 1$. Our resulting state of confidence is shown in Figure 5.3a as curve (1). We see already an immediate shift towards lower values of A. The evidence $k = 0$, $m = 1$, is already sufficient to tell us that A cannot have value 1.0.

Now suppose we run another trial and again x_N is less than $\hat{x}_N$. Our evidence is now $k = 0$, $m = 2$, which from equation (5.12) yields curve (2) in Figure 5.3a. We see a substantial further shift to the left.

Continuing in this way, suppose in curve (3) that the evidence is zero hits out of five trials, in curve (4) zero out of ten trials and so on. As the number of trials increases with no values $x_N > \hat{x}_N$, our confidence increases that A is a small number. This is shown by the increasing weight of the probability curves on the lower end of the A scale. Figure 5.3b, which shows the cumulative probability functions calculated by equation (5-13) for the curves in Figure 5.3a, also demonstrates that the probability that A is a large number decreases as the number of trials increases.

Figures 5.4a and 5.4b show similar results for a series of runs in which 30 percent of the trials have $x_N > \hat{x}_N$. Our best estimate in all cases is that $A = 0.3$. Our confidence in that estimate increases with the number of trials, as shown by the increasing narrowness of the curves in Figure 5.4a.

FIGURE 5·3a

POSTERIOR PROBABILITY DENSITY FUNCTIONS OF A FOR VARIOUS SAMPLE SIZES m, GIVEN k/m = 0

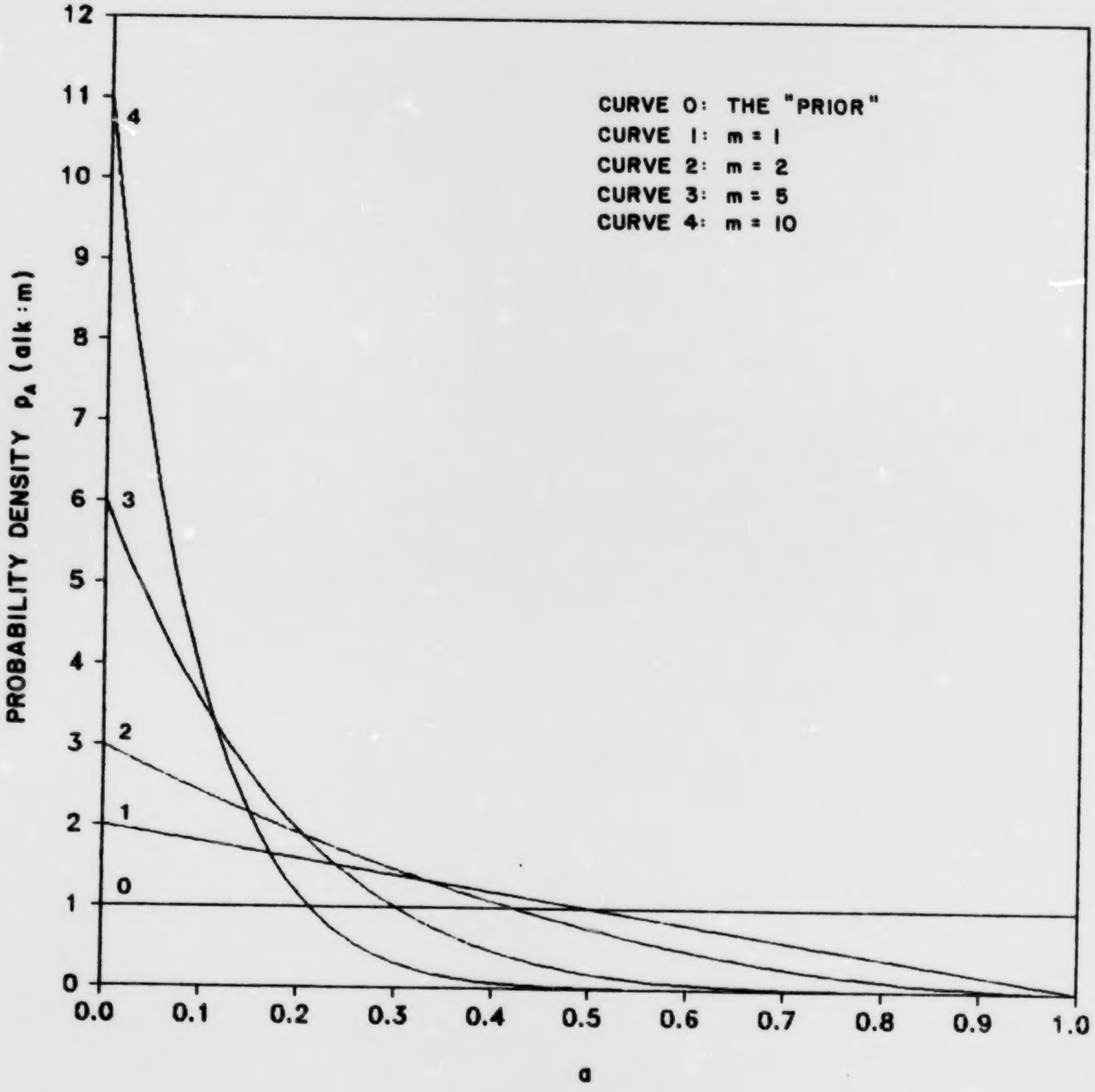


FIGURE 5·3b

CUMULATIVE PROBABILITY FUNCTIONS OF A FOR VARIOUS SAMPLE SIZES m, GIVEN k/m = 0

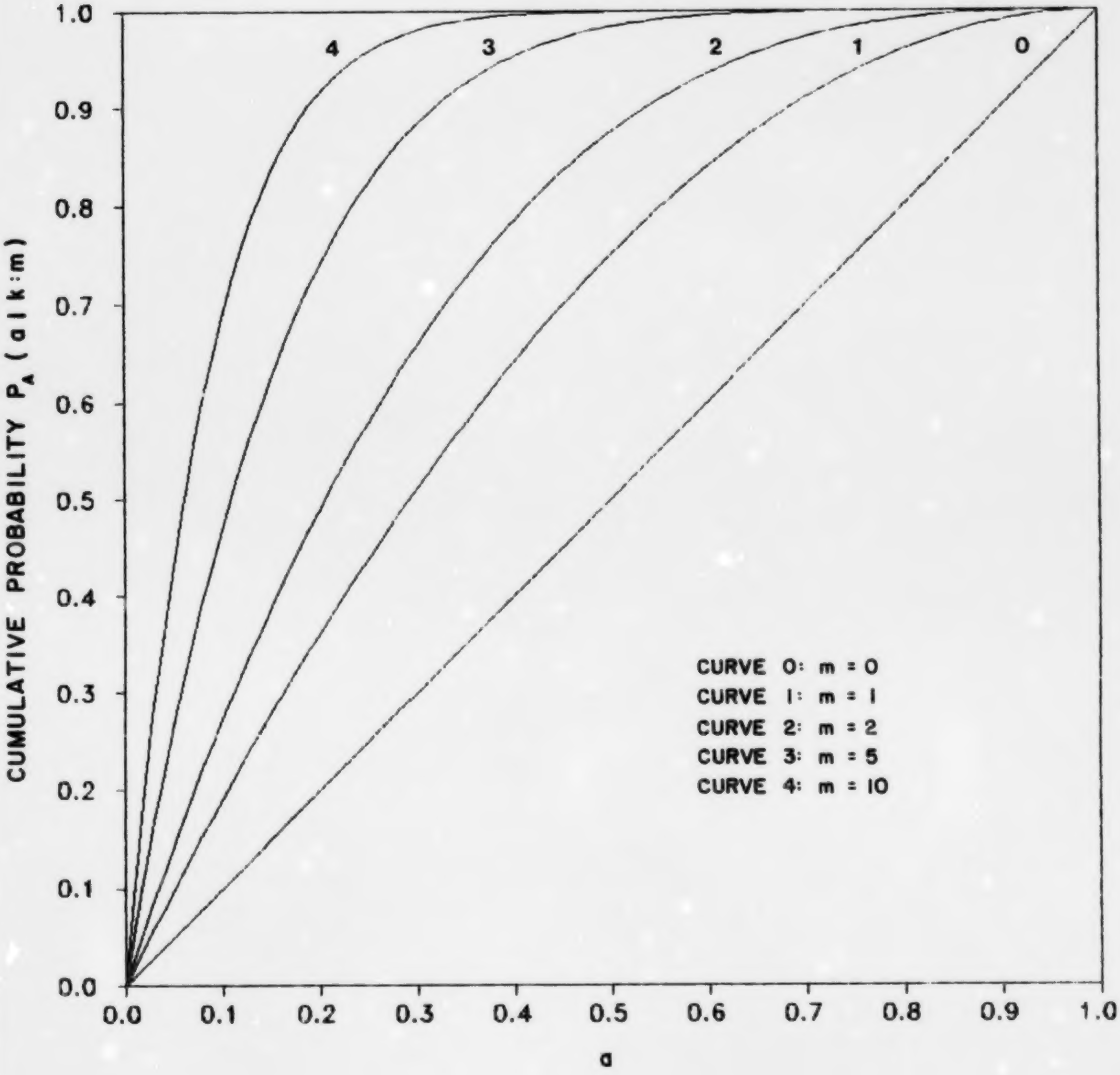


FIGURE 5·4a

POSTERIOR PROBABILITY DENSITY FUNCTIONS OF A FOR VARIOUS SAMPLE SIZES m, GIVEN k/m = 0·3

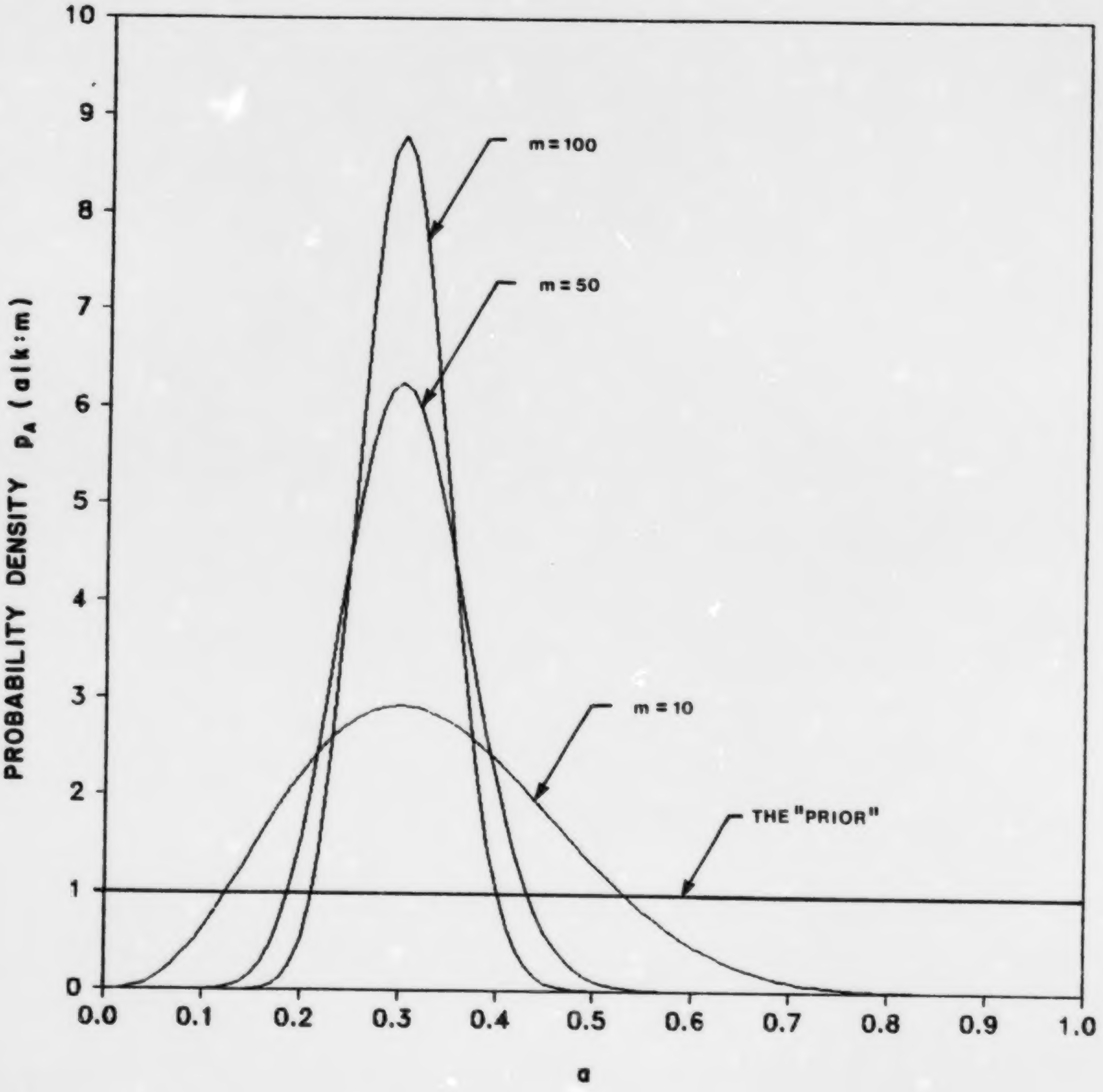
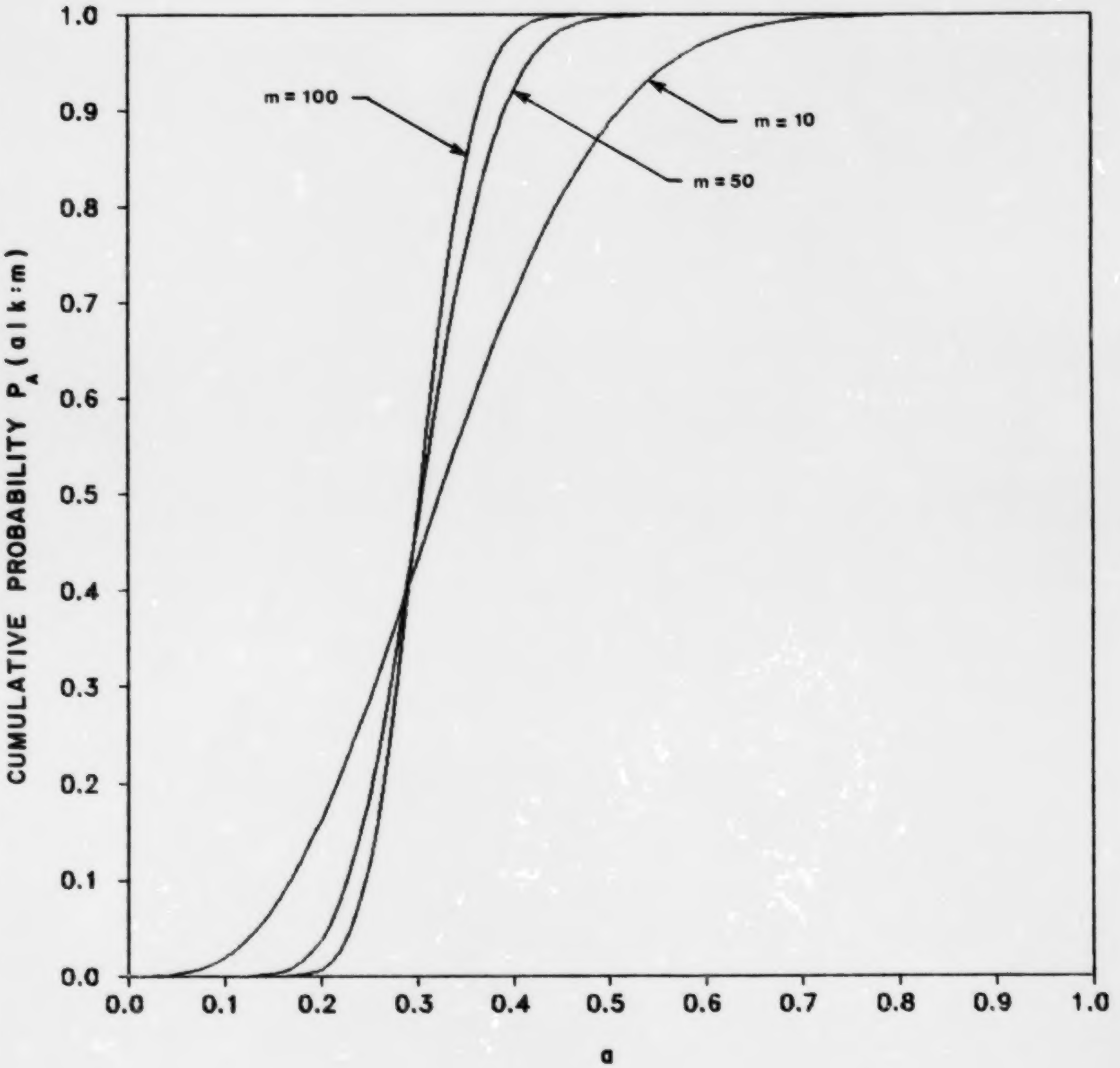


FIGURE 5·4b

CUMULATIVE PROBABILITY FUNCTIONS OF A FOR VARIOUS SAMPLE SIZES m, GIVEN k/m = 0·3



Figures 5.5a, 5.5b, 5.6a and 5.6b show the results of similar series of calculations for the cases $k/m = 0.1$ and $k/m = 0.05$. (Note the expanded scale on the A axis in Figure 5.6a relative to the A axis in Figure 5.5a).

These curves illustrate the computational difficulty mentioned in Section 4.1. If the regulator desires to require that:

> "The area A, exceeding $\hat{x}_N$, shall be no greater than $\hat{A}$."

then, since there is computational uncertainty (shimmer) in the calculation of A, the regulation must be put into the even more cumbersome form:

> "With confidence level $\hat{c}$ or greater, the area A, exceeding $\hat{x}_N$, shall be no greater than $\hat{A}$."

Figures 5.3 to 5.6 show that the smaller $\hat{A}$ is, the greater the number of trials required to reach a given level of confidence, c, that A is less than $\hat{A}$. But even if we do run a large number of trials, we are still addressing only the shimmer effects. The above procedure does not evaluate the inherent non-robustness in A resulting from the non-robustness of the tails of the input distributions. Hence, we return to our earlier statement that the decision maker must learn to consider the whole curve $p(x_N)$ rather than focus on particular values of x_N and A.

FIGURE 5·5a

POSTERIOR PROBABILITY DENSITY FUNCTIONS OF A FOR VARIOUS SAMPLE SIZES m, GIVEN k/m = 0·1

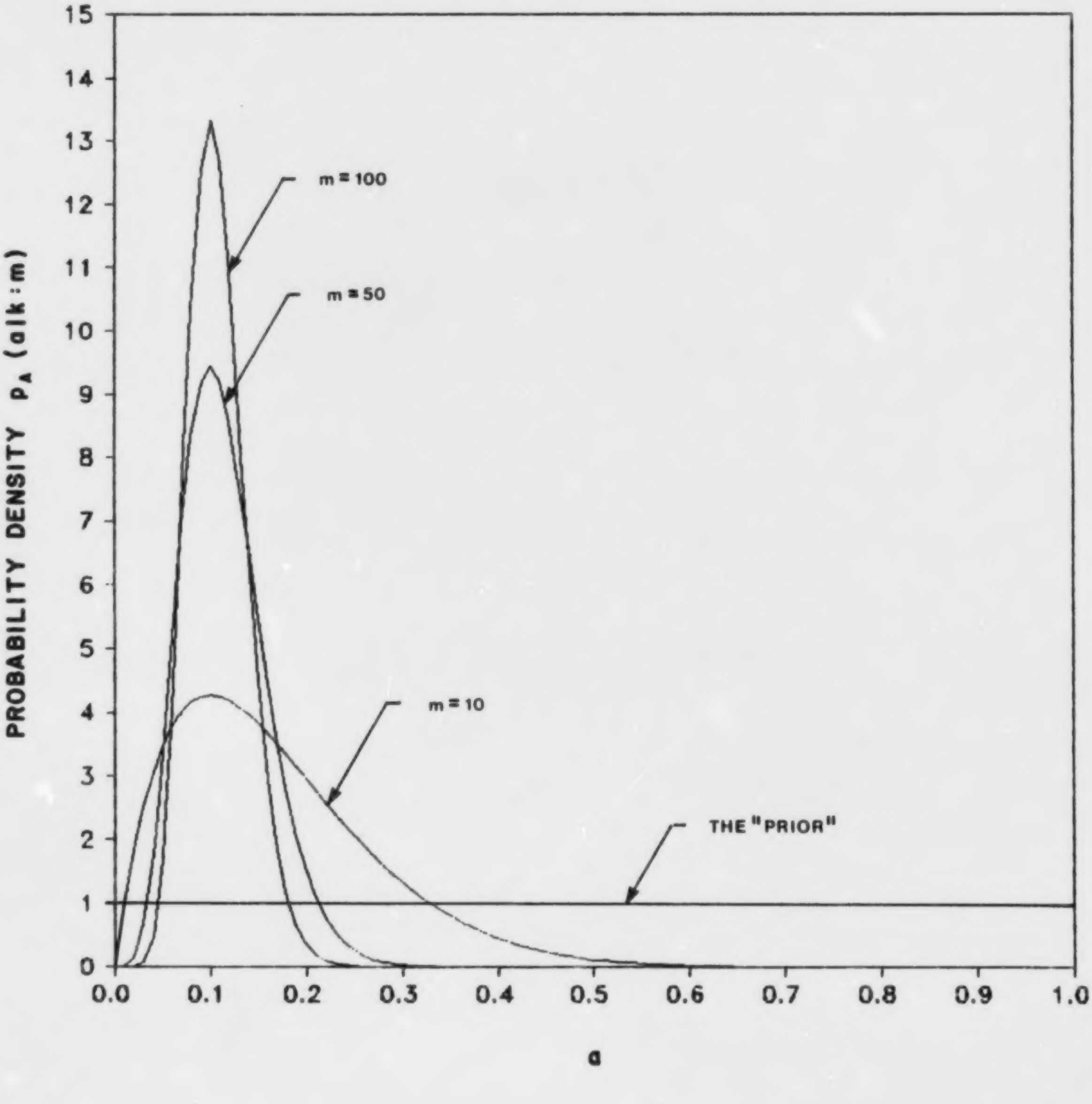


FIGURE 5·5b

CUMULATIVE PROBABILITY FUNCTIONS OF A FOR VARIOUS SAMPLE SIZES m, GIVEN k/m = 0·1

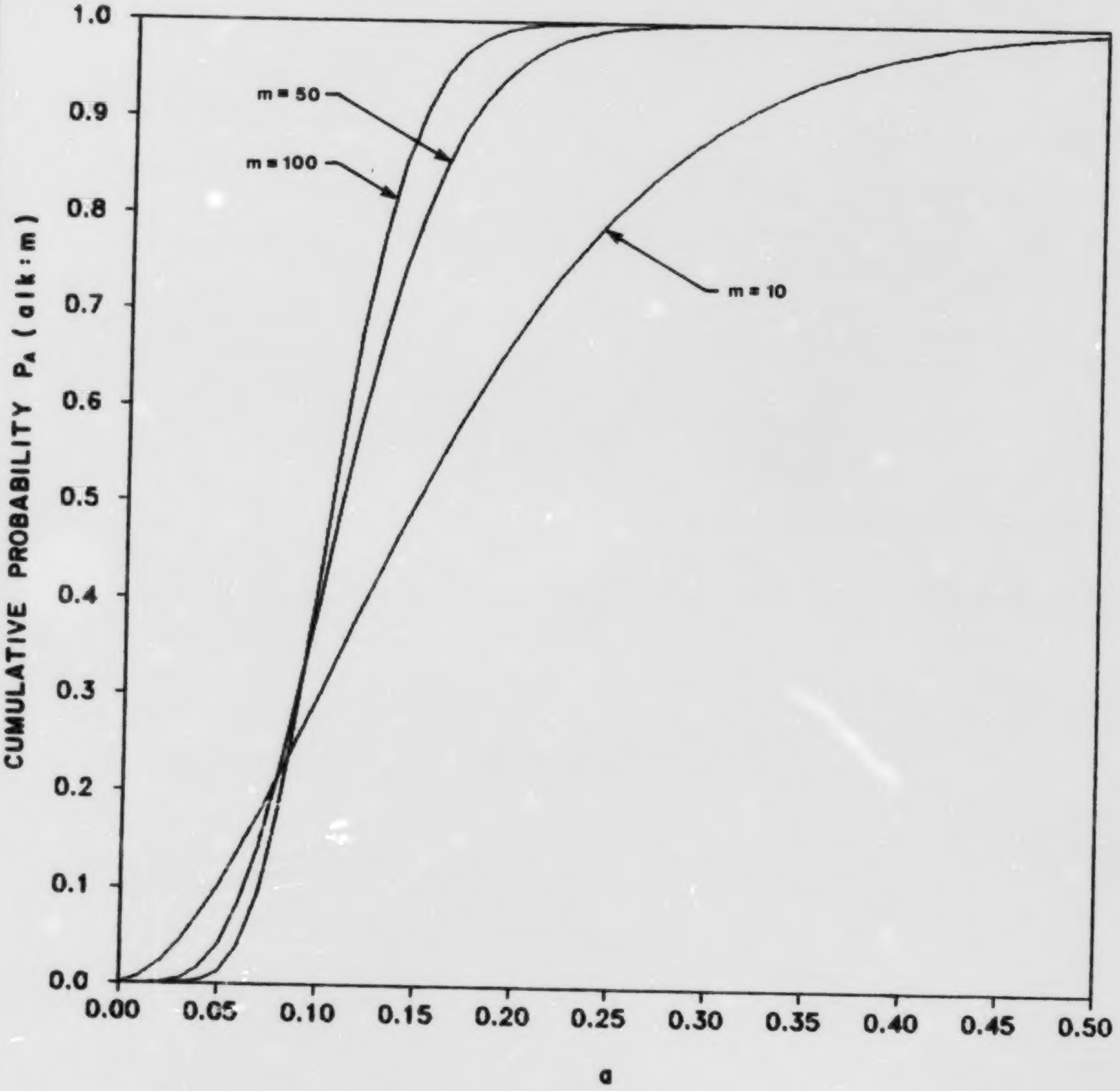


FIGURE 5·6a

POSTERIOR PROBABILITY DENSITY FUNCTIONS OF A FOR VARIOUS SAMPLE SIZES m, GIVEN k/m = 0·05

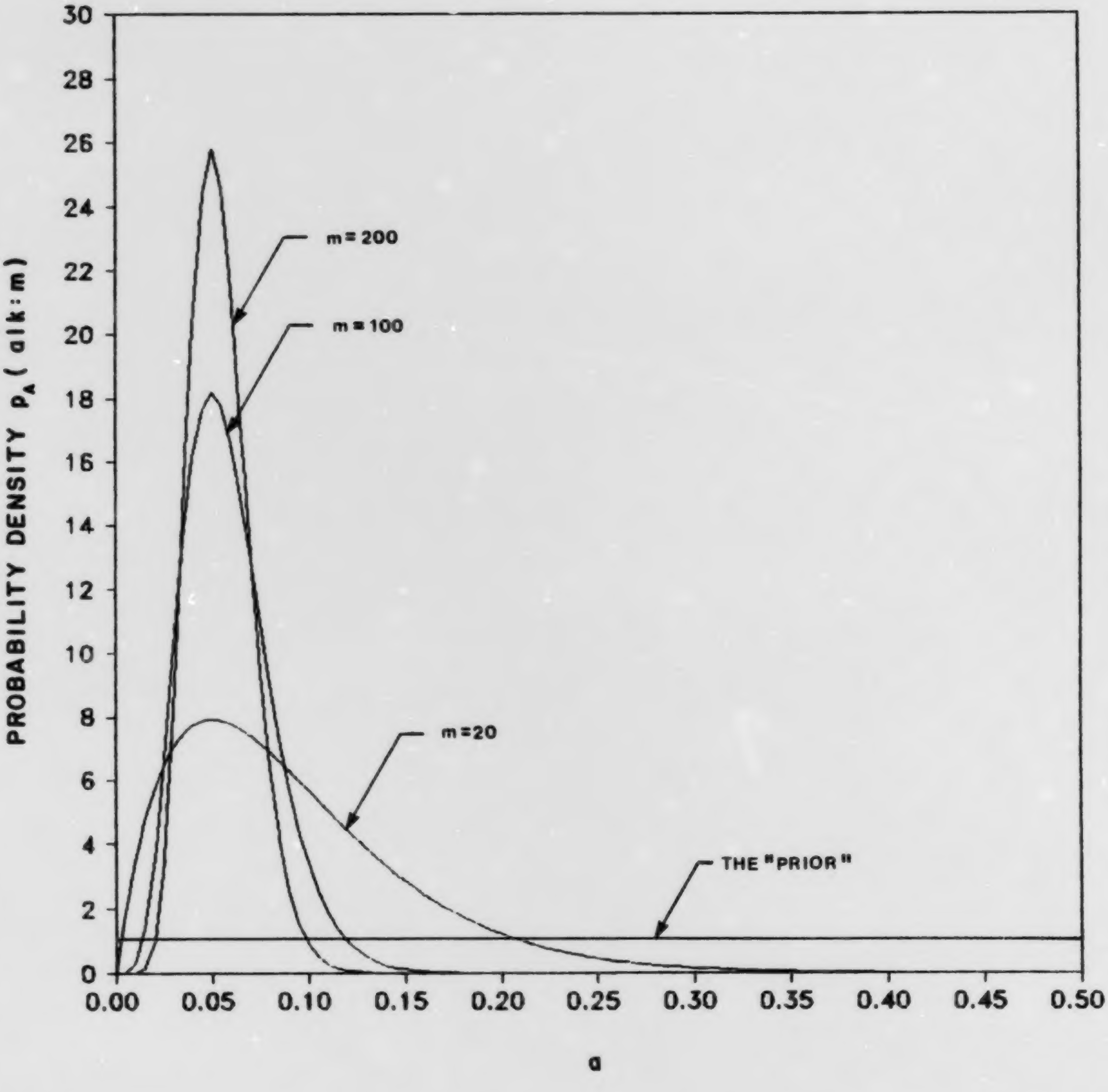
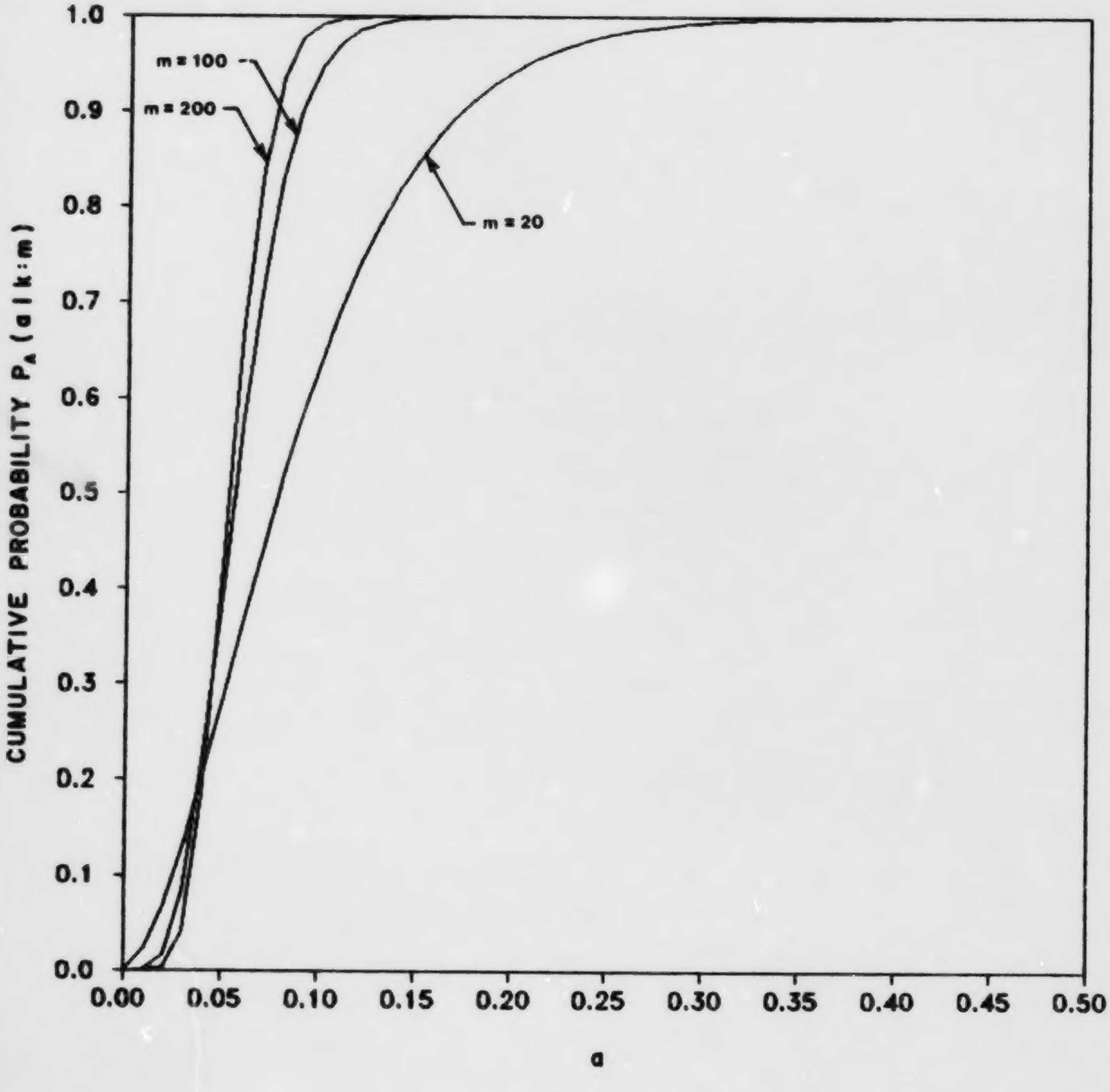


FIGURE 5·6b

CUMULATIVE PROBABILITY FUNCTIONS OF A FOR VARIOUS SAMPLE SIZES m, GIVEN k/m = 0·05



REFERENCES

Andres, T.H., 1984a. Personal Communication from Terry H. Andres to D.B. Chambers, August 1984 concerning concept of "shimmer".

Andres, T.H., 1984b. "Confidence Bounds on an Empirical Distribution" Atomic Energy of Canada Draft document, Whiteshell Nuclear Research Establishment, February.

Kaplan, S., 1983. "On a 'Two-Stage' Bayesian Procedure for Determining Failure from Experimental Data," PLG-0191, IEEE Transactions on Power Apparatus and Systems, Vol. PAS-102, No. 1, January.

Kaplan, S., 1982. "On Safety Goals and Related Questions," Reliability Engineering, Vol. 3, pp. 267-277.

Kaplan, S., 1981. "On the Method of Discrete Probability Distributions in Risk and Reliability Calculations - Application to Seismic Risk Assessment," Risk Analysis, Vol. 1, No. 3, pp. 189-196, September.

Pearson, K., 1948. "Tables of the Incomplete Beta Function." Biometrika Office, University College, London.

Rinehart and Company, 1965. Mathematical Tables, New York, March, p. 267.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada - Phase I." Research Report prepared for the National Uranium Tailings Program, CANMET, EMR, Under Supply and Services Canada Contract Number OSQ83-00236.

DISCUSSION PAPER ON
PROBABILITIES OF INDIVIDUAL CELLS OR TAIL AREA UNDER MONTE CARLO SAMPLING

PREPARED BY

PARK M. REILLY
WATERLOO, ONTARIO

FOR
SENES CONSULTANTS LIMITED
WILLOWDALE, ONTARIO
June 1985

Abstract

This paper discusses the relative merits of latin hypercube sampling (LHS) and simple random sampling (RS). It is concluded that the use of LHS will normally produce a mean with smaller variance than that produced using RS. On the other hand it is noted that there are exceptions to the foregoing observation and that LHS is more awkward to implement on a computer. The paper also discusses the application of a Bayesian approach to estimating the probability that the value of a variable estimated using Monte Carlo techniques will fall below or exceed a stated value.

TABLE OF CONTENTS

		Page
1.0	BACKGROUND	1
2.0	LATIN HYPERCUBE SAMPLING	1
3.0	ESTIMATING THE PROBABILITY THAT THE TRUE VALUE OF A VARIABLE LIES IN A GIVEN REGION	2
REFERENCES		7

1.0 BACKGROUND

In most ways this is a reworking of the excellent discussion papers and past presentations by others, especially Kaplan and Associates, Inc. and Dr. Terry Andres. While this author disagrees with some of their positions in minor detail, my areas of agreement with them are much broader and more important. I plan to discuss a few matters in order mostly to provide a different viewpoint.

2.0 LATIN HYPERCUBE SAMPLING

I must admit that I have been metaphorically dragged kicking and screaming into examining the merits of Latin hypercube sampling (LHS) and have emerged a mildly enthusiastic supporter. In every practical example which I have seen, when used for the purpose of estimating the mean of an output variable, the LHS method has given a smaller variance of estimate than simple random sampling (SRS); in the simple and therefore common cases the variance is very much better. In some contrived examples which others and I have constructed, LHS has given results which are only slightly better than SRS and indeed in one case suggested by Dr. Terry Andres SRS gives results slightly better than LHS.

I concede that if one were required to determine in advance exactly how many trials would be made, arguments can be made against LHS. However, because of its usual advantage in variance, I believe that LHS is ordinarily to be preferred in determining the mean. If after an initial run, the mean has not been found with sufficiently small estimated variance, a further run or runs may be made and the results pooled.

On the other hand, computer coding of LHS is much more awkward and much more memory is used. The full cost-benefit analysis depends on the individual computer installation and it is quite possible to imagine situations where SRS is cheaper and more convenient overall. This probably applies particularly on smaller jobs.

Finally, the estimation of a distribution function of any kind reduces to the problem of estimating the mean number of counts in one or more cells. The remarks made above apply as well to this problem as to that of estimating the mean of an output variable.

3.0 ESTIMATING THE PROBABILITY THAT THE TRUE VALUE OF A VARIABLE LIES IN A GIVEN REGION

It is a very important problem to estimate the probability that a certain variable will exceed or fall below a stated value. Sometimes it is also important to know how probable it is that a variable will fall between two stated limits.

In discussing this subject I want to emphasize that by probability I mean strength of belief which you hold or are entitled to hold about a proposition. It can be measured by making a subjective comparison with the outcome of some mechanical operation such as rolling dice, etc., the long term relative frequency of which is known or calculable.

This leads to a difficulty with Monte Carlo calculations. Typically we enter probability distributions of at least some of the input variables. Probability distributions rather than constants are needed because of our uncertainties about the values of the variables. The uncertainty may be of the form that quantifies our information about a number which is truly a constant but is unknown, or possibly it is quite a knowable quantity like wind speed, but which is variable. In either case, and in the many intermediate ones, we enter values of input variables in the form of probability distributions. As far as our uncertainty is concerned, the probability distributions have the same meaning in all cases.

After making Monte Carlo trials we produce histograms of output variables. If enough trials, say a billion, were made, the output histograms would be practically true probability distributions of the same nature as the input ones. With a typically small number of trials, however, there is doubt as to true probability because of "raggedness", or "shimmer" in the histogram. Hence we can ask the question, "Given a Monte Carlo run with a stated number of trials, which resulted in a stated number of counts in a certain interval on the range of an output variable, how strongly may we believe (i.e. what is the probability) that the value of the variable is in that interval?"

It is convenient and usually realistic to assume that only a small fraction of the total counts are expected to fall in the interval. Let the long term relative frequency (which is here numerically equal to what I have defined as probability) of a Monte Carlo count entering the interval be p. If the total number of trials is N, the mean (expected) number of counts in the interval is Np. Since each Monte Carlo trial is a Bernoulli trial with respect to a count entering the interval, the number of counts is a binomial variable and therefore has variance Np(1-p). Now if p is small, this is nearly the same as Np and hence under this condition the number of counts may be considered a Poisson random variable with parameter $\lambda = Np$, because the Poisson probability law describes the number of successes in Bernoulli trials under conditions where the mean and variance of the number of successes are equal and the number of trials is large.

Let us now adopt a Bayesian approach and attempt to estimate , given that we have observed n counts which fell in the interval of interest out of a total of N Monte Carlo trials.

A consideration in the use of Bayesian methods is the choice of prior distribution for the parameter λ. It is desired (in most cases) to express as much prior ignorance as possible. There is no complete agreement as to how best to do this but almost all authorities either agree with Jeffreys (1961) and Lindley (1965) in assuming the prior probability density function to be proportional to λ^{-1} or agree with Box and Tiao (1972) in assuming it proportional to $\lambda^{-1/2}$. With any reasonable amount of data it makes little difference which is chosen and here the former will be used. Parenthetically, similar recommendations, i.e. $p^{-1}(1-p)^{-1}$ and $p^{-1/2}(1-p)^{-1/2}$, are made for the case of the binomial parameter p.

The likelihood function as always is the probability function, here Poisson, with the data substituted numerically and therefore it is a function of the parameter. It is

$$\frac{\lambda^{n}e^{-\lambda}}{n!}$$

By Bayes' theorem the probability distribution of the parameter λ given the data is proportional to the product of the prior probability density and the likelihood, i.e.

$$Df(\lambda) \alpha \quad \lambda^{n-1}e^{-\lambda}$$

where Df means "probability density function" and n! has been omitted because it is constant.

This is recognizable as the variable part of a gamma probability density function. Supplying the normalizing constant the probability density function is

$$Df(\lambda) = \frac{\lambda^{n-1} e^{-\lambda}}{(n-1)!}$$

It is of more interest to make the transformation to the probability density function for p. The result is

$$Df(p) = \frac{N^n}{(n-1)!} p^{n-1} e^{-pN}$$

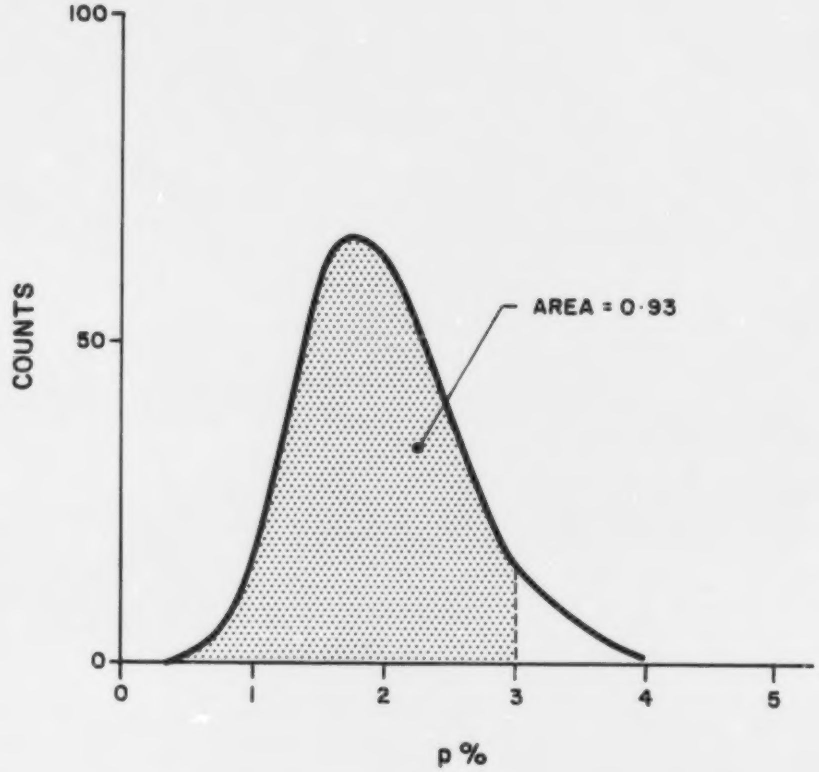
It is of interest and of considerable practical value to make a further transformation to $\omega = 2Np$. It is found that this quantity is distributed as χ^2 with 2n degrees of freedom. As an example of this, suppose that 500 Monte Carlo trials have been made and 10 of them exceeded a certain regulated value. What is the probability (how strongly are you entitled to feel) that the true probability of exceeding this regulation is less than 3%? Here $\omega = 2 \times 500 \times .03 = 30$. From tables, the probability that χ^2 on 2 x 10 =20 degrees of freedom is less than 30 is 93.0% and consequently we may feel 93% strongly that the probability of exceeding the regulation is less than 3% (see Figure 1).

This type of calculation is valuable because it gives the integral of the probability density function for p. In order to visualize the probability distribution for p, it is easy to plot the probability density function as given. This is done in Figure 1 for N = 500 and n = 10 as in the previous example.

Suppose N = 500 and n = 10, and you want a single number to believe in. What is it? This can be justified as a best guess (not necessarily a good one) by noting that there is an infinite set p_1, p_2, ... that you might believe in, each having probability represented by the infinitesimal area under the probability density function for p at that value. In the long run, which value is the best one? Obviously the value found by integrating pDf(p) over all possible p. This is the definition of the mean of p and it turns out to be n/N, a not surprising value! In this case it is .02 or 2%.

FIGURE 1

PROBABILITY DENSITY FUNCTION FOR p, WITH $N^{*}=500$ AND $n^{**}=10$



NOTE:

* N IS THE NUMBER OF MONTE CARLO TRIALS

** n IS THE NUMBER OF MONTE CARLO TRIALS IN INTERVAL (>3)

REFERENCES

Box, G.E.P., and G.C., Tiao, 1972. "Bayesian Inference in Statistical Analysis." Addison-Wesley.

Jeffreys, H., 1961. "Theory of Probability." 3rd Edition, Oxford: Clarendon Press.

Lindley, D.V., 1965. "Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 2 - Inference." Cambridge University Press.

DISCUSSION PAPER
ON
CONFIDENCE REGIONS FOR OUTPUT DISTRIBUTIONS FROM MONTE CARLO CALCULATIONS

Prepared by

Park M. Reilly and T.A. Duever
Waterloo, Ontario

For

SENES Consultants Limited
Willowdale, Ontario

April 1985

ABSTRACT

This paper describes a method of stating bounds within which the true probability density function of the output variable of a Monte Carlo calculation may be expected to lie, at a given confidence level. The method is an extension of work by Cheng and Iles and by Scheffé in which a probability density function of known mathematical form is fitted to observed relative frequency values and, using conventional statistical methods, a joint confidence region for the parameter values is found. The applicability of the method to output results of environmental assessment models is demonstrated. The confidence bands calculated using this method were found to be narrower than the bands derived using a procedure developed by Kolmogorov and Smirnov.

Table of Contents

Page

1.0 INTRODUCTION AND BACKGROUND 1

2.0 GENERAL FORMS FOR OUTPUT PROBABILITY DENSITY FUNCTIONS 2

3.0 FITTING THE PROBABILITY DENSITY FUNCTION TO THE DATA 4

4.0 ESTABLISHING CONFIDENCE BOUNDS FOR THE FITTED FREQUENCY FUNCTION 6

5.0 ESTABLISHING CONFIDENCE BOUNDS FOR THE CUMULATIVE PROBABILITY FUNCTION 7

6.0 EXAMPLE OF CONFIDENCE BOUNDS ESTIMATION 9

7.0 CONCLUSIONS 11

REFERENCES 12

APPENDIX A: NEWTON-RAPHSON ITERATIVE SOLUTION 13

APPENDIX B: ANALYTICAL SOLUTION FOR $r = 1$ 14

APPENDIX C; KOLMOGOROV CONFIDENCE BANDS 15

List of Figures

Follows Page

Figure 1 Fitted Probability Density Functions For s=2, s=4 and s=6 — 9

Figure 2 95% Confidence Limits for Probability Density Function, Classroom Economic Study (s=4) — 10

Figure 3 95% Confidence Limits for Probability Density Function, DOSRN 100 Trials - 2 Parameter Model — 10

Figure 4 95% Confidence Limits for Probability Density Function, DOSRN 500 Trials - 2 Parameter Model — 10

Figure 5 95% Confidence Limits for Cumulative Distribution Function, DOSRN 100 Trials - 2 Parameter Model — 11

Figure 6 95% Confidence Limits for Cumulative Distribution Function, DOSRN 500 Trials - 2 Parameter Model — 11

List of Tables

Follows Page

Table 1 Illustration of Cheng and Iles Technique — 6

Table 2 Variable DOSRN 100 Trials — 10

Table 3 Variable DOSRN 500 Trials — 10

1.0 INTRODUCTION AND BACKGROUND

The most important output from a Monte Carlo calculation is often the probability density function for one or more output variables. Each probability density function is obtained in approximate form as a set of counts in cells which correspond to different values of the variable. The observed relative frequencies in the cells provide an approximation to the probability density function of the variable.

If the number of Monte Carlo trials were extremely large (say 10^9) and the number of cells moderate (say 20), it would be expected that the graph of the relative frequency values versus the values of the variable corresponding to different cells would be a smooth curve which would agree very closely with the true probability density function. It would also vary little from one Monte Carlo run with the same number of trials to the next.

If however only 200 Monte Carlo trials are made with 20 cells, the relative frequency diagram can be expected to be "ragged", to vary a good deal from one run of 200 trials to the next, and the smooth curve drawn through the relative frequency points cannot be expected to agree closely with the true probability density function. It is the object of this exposition to describe a method of stating bounds within which the true probability density function may be expected to lie, at a given confidence level.

In the Phase I SENES study (SENES 1984) a distribution-free technique based on an extension of the Kolmogorov-Smirnov goodness of fit test (Kolmogorov 1941), as described for example in Kendall and Stuart (1961), was used to develop confidence bands for the cumulative distribution function. The test is based on the maximum absolute deviation between the empirical (i.e. observed) cumulative distribution function, and the underlying true cumulative distribution. Unfortunately, while the method is attractive because of its simplicity, it does have the disadvantage of being too conservative, particularly in the tails of the distribution. Usually, ordinate values of less than zero for the left tail and greater than one for the right tail are obtained. In practice the confidence bands are truncated at zero or one respectively. This consideration prompted the further investigation of techniques for establishing confidence bands which led to the present paper.

The method discussed in this paper is an extension of the approach of Cheng and Iles (1983) and of Scheffé (1959), in which a probability density function of known mathematical form is fitted to the observed relative frequency values by estimating the values of the parameters in the probability density function. At the same time, by conventional statistical methods, a joint confidence region for the parameter values is found. Then the set of mathematical functions which might, at the confidence level being used, include the true probability density function is the set of mathematical functions obtained by substituting into the known mathematical form of the probability density function all values of the parameters within their joint confidence region. The envelope of this set of functional relations is the confidence region, at the confidence level being used, for the probability density function of the variable.

2.0 GENERAL FORMS FOR OUTPUT PROBABILITY DENSITY FUNCTIONS

The first requirement in calculating the confidence interval for the probability density function is a method of establishing flexible mathematical forms for the probability density functions to which the observed data may be fitted. Occasionally the output probability density function is of known form because of known forms for the input variables and a simple function relating them to the output variable. In the more interesting cases however, the probability density functions for the output variables are not of simple form.

The following form is proposed for the output probability density function of a random variable represented, possibly in transformed form, by x. The cell mid-points are represented by $x_1, x_2, \ldots, x_m$.

$$p(x) = \exp\,(\Theta_o + \Theta_1 x + \Theta_2 x^2 + \ldots\ldots + \Theta_s x^s) \qquad (1)$$

$$\sum_{i=1}^{m} p(x_i) = 1$$

where p is the probability density function, s is an integer representing the degree of the approximating polynomial, and m is the number of cells.

The variable Θ_o is an auxiliary variable which can be interpreted as a normalizing constant. It can be eliminated if desired by writing (1) in the following form:

$$p(x) = \frac{\exp\,(\Theta_1 x + \Theta_2 x^2 + \ldots\ldots\ \Theta_s x^s)}{\sum\limits_{i=1}^{m} \left[\exp\,(\Theta_1 x + \Theta_2 x^2 + \ldots + \Theta_s x^s)\right]} \qquad (2)$$

which obviously sum to unity over all x_i

In actual practice the data are given as cell counts at a number, usually 10 to 20, of cells (bins), each cell usually representing a small range of values of x. In that case, p(x) in equation (1) can be interpreted as the probability of a count entering the cell which corresponds to the value of x at the mid-point of the cell. It will be only approximately true that if the p(x), interpreted as cell probabilities, sum to 1, then p(x) interpreted as a probability density as was done above, will integrate to 1. If necessary this deficiency can be corrected by adjusting the value of the normalizing constant.

3.0 FITTING THE PROBABILITY DENSITY FUNCTION TO THE DATA

Suppose that cell counts have been observed in m cells and the ith cell, the mid-point of which corresponds to the value x_i of the random variable, has recorded n_i counts. It is desired to estimate Θ_j, j=0, 1, ..., s so as to fit equation (1). At this time $p(x_i)$ in equation (1) will be interpreted as the probability of a count entering the ith cell.

Such a situation is described by the multinomial likelihood function:

$$L(\Theta_1, \ldots \Theta_s) = \prod_{i=1}^{m} [p_i(x_i)]^{n_i} \tag{3}$$

A very good estimate of $\underline{\Theta}$, the maximum likelihood estimate, is obtained by maximizing L, or more conveniently by minimizing its negative logarithm:

$$S(\underline{\Theta}) = -\ln L(\underline{\Theta}) = -\sum_{i=1}^{m} n_i \ln p_i \tag{4}$$

$$\ln p_i = \Theta_o + \sum_{k=1}^{s} \Theta_k x_i^k \tag{5}$$

$$\text{and} \quad \sum_{i=1}^{m} p_i = 1 \tag{6}$$

This system can be solved by first differentiating equation (6):

$$\frac{\delta}{\delta\,\Theta_j} \sum_{i=1}^{m} \exp\left\{ \Theta_o + \sum_{k=1}^{s} \Theta_k x_i{}^k \right\} = 0$$

$$\sum_{i=1}^{m} p_i \left(\frac{\delta\Theta_o}{\delta\Theta_j} + x_i{}^j \right) = 0$$

Therefore

$$\frac{\delta\Theta_o}{\delta\Theta_j} = - \frac{\sum_{i=1}^{m} p_i x_i^j}{\sum_{i=1}^{m} p_i} = - \sum_{i=1}^{m} p_i x_i{}^j; \quad j = 1, \ldots, s \qquad (7)$$

To find the minimum of equation (4) differentiate with respect to $\Theta_j; j=1,2,\ldots,s$:

$$\frac{\delta S}{\delta\Theta_j} = - \sum_{i=1}^{m} n_i \left[\frac{\delta\Theta_o}{\delta\Theta_j} + x_i{}^j \right]$$

$$= -N \frac{\delta\Theta_o}{\delta\Theta_j} - \sum_{i=1}^{m} n_i x_i{}^j \qquad (8)$$

where

$$N = \sum_{i=1}^{m} n_i$$

Substituting equation (7) into equation (8) and setting equation (8) to zero:

$$N \sum_{i=1}^{m} p_i x_i{}^j = \sum_{i=1}^{m} n_i x_i{}^j \qquad j = 1,2,\ldots,s \qquad (9)$$

Note that for j=0 equation (9) reduces to the constraint from which Θ_o is calculated. While there is a theoretical distincion made between Θ_o and $\underline{\Theta}$, the vector of all Θ's except Θ_o, they are all found in practice by solving the following nonlinear system of (s+1) equations in (s+1) unknowns.

$$N \sum_{i=1}^{m} p_i x_i^j = \sum_{i=1}^{m} n_i x_i^j; \qquad j = 0,1,\ldots,s \tag{10}$$

Equation (10) can be solved by the Newton-Raphson method as described in Appendix A.

4.0 ESTABLISHING CONFIDENCE BOUNDS FOR THE FITTED FREQUENCY FUNCTION

The probability function having been fitted to each observed cell frequency, the approach of Cheng and Iles permits confidence bounds on it to be established as outlined previously. The first step in this is to establish a joint confidence region for the parameters $\underline{\Theta}$. This region is:

$$(\underline{\Theta} - \underline{\hat{\Theta}})\underline{\acute{M}}(\underline{\Theta} - \underline{\hat{\Theta}}) \leq \gamma \tag{11}$$

where $\underline{\Theta}$ is the sx1 vector of possible parameter values, $\underline{\hat{\Theta}}$ is the sx1 vector of parameter estimates and $\underline{M}$ is the sxs Fisher information matrix for the parameters. Also γ is a number which has approximately the value of a χ^2-distributed random variable with s degrees of freedom and is exceeded with probability (1-a), where a is the desired confidence level. The (j,k)th element of the sxs Fisher information matrix is given by:

$$\underline{M} = \left[N \sum_{i=1}^{m} p_i x_i^{j+k} - N\left(\sum_{i=1}^{m} p_i x_i^{j}\right)\left(\sum_{i=1}^{m} p_i x_i^{k}\right) \right] \tag{12}$$

The next step in the Cheng and Iles procedure is to choose values of $\underline{\Theta}$, so as to maximize and minimize equation (1) subject to equation (11). Since exponentiation is a monotone function it will be sufficient to maximize and minimize lnp(x). The required Lagrangian is:

$$L(\underline{\Theta}) = \Theta_o + \sum_{k=1}^{s} \Theta_k x^k - \frac{1}{2} \lambda (\underline{\Theta}-\hat{\underline{\Theta}})' \underline{M} (\underline{\Theta}-\hat{\underline{\Theta}}) \qquad (13)$$

Solving this leads to the following equation for calculating values of $\underline{\Theta}$ which maximize and minimize p:

$$\underline{\Theta} = \hat{\underline{\Theta}} \pm \sqrt{\frac{\gamma}{(\underline{c}-\underline{b})'\underline{M}^{-1}(\underline{c}-\underline{b})}} \; \underline{M}^{-1} (\underline{c}-\underline{b}) \qquad (14)$$

where

$\underline{b} = [\sum_{i=1}^{m} p_i x_i{}^j]$ jth element of the sxl vector $\underline{b}$

$\underline{c} = [x^j]$ jth element of the sxl vector $\underline{c}$

Note that equation (14) requires an iterative solution since $\underline{b}$ is a function of $\underline{\Theta}$. A scheme based on simple iteration was found to give rapid convergence.

Having found the limiting values of $\underline{\Theta}$, the limits on p are found by substituting these in equation (1).

5.0 ESTABLISHING CONFIDENCE BOUNDS FOR THE CUMULATIVE PROBABILITY FUNCTION

If equation (1) is interpreted as describing the probability density function for X, the cumulative probability function is defined by:

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} p(u)du = \int_{-\infty}^{x} \exp\left\{\Theta_o + \sum_{k=1}^{s} \Theta_k u^k\right\} du \qquad (15)$$

Confidence limits for F(x) for a fixed x may be found as in the last section by maximizing and minimizing F(x) by varying $\underline{\Theta}$ within the limits given by equation (11). They are found in terms of the vectors $\underline{\Theta}$ which produce the maximum and minimum for F(x). The limiting $\underline{\Theta}$ are found from the following equation, applied iteratively:

$$\underline{\Theta} = \hat{\underline{\Theta}} \pm \sqrt{\frac{\gamma}{\underline{r}' \underline{M}^{-1} \underline{r}}}\, \underline{M}^{-1} \underline{r} \qquad (16)$$

where $\underline{\Theta}$, $\hat{\underline{\Theta}}$, γ and $\underline{M}$ are as defined previously and $\underline{r}$ is the sx1 vector:

$$\underline{r} = \int_{-\infty}^{x} [\underline{d}+\underline{e}] \exp\left\{\Theta_o + \sum_{k=1}^{s} \Theta_k u^k\right\} du \qquad (17)$$

where

$$\underline{d} = \frac{\delta\Theta_o}{\delta\underline{\Theta}} = -\int_{-\infty}^{\infty} u^j \exp\left\{\Theta_o + \sum_{k=1}^{s} \Theta_k u^k\right\} du$$

is the jth element of the sx1 vector $\underline{d}$

$\underline{e} = [u^j]$ is the jth element of the sx1 vector $\underline{e}$.

Note that Θ_o is now given by:

$$\Theta_o = -\ln\left[\int_{-\infty}^{\infty} \exp\left\{\sum_{k=1}^{s} \Theta_k u^k\right\} du\right] \qquad (18)$$

For the case where s=1, the required integral can be found analytically as shown in Appendix B. For s>1, numerical integration will be required. In this paper we have chosen to work with the probability distribution in discrete form. This has the advantage of eliminating the need for numerical integration.

When the limit producing values of $\underline{\theta}$ have been found, the bounds on F(x) are found by substitution into the integral in equation (15).

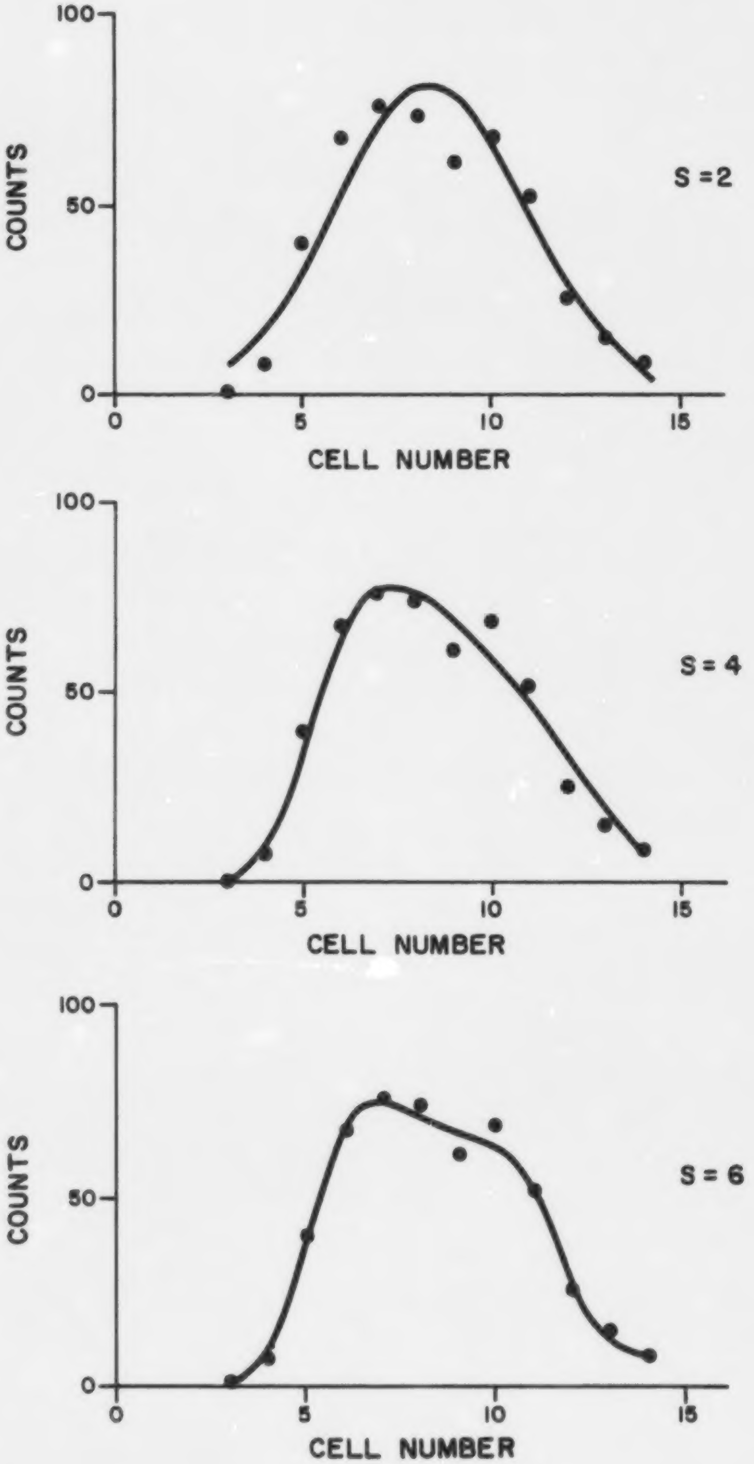
6.0 EXAMPLE OF CONFIDENCE BOUNDS ESTIMATION

Table 1 shows cell counts obtained in N=500 trials in a classroom example which used a random sampling program (OPANAL) in an economic study. The data were fitted using s=2, s=4 and s=6. Using the standard χ^2 test for goodness-of-fit the χ^2 values are respectively 26.54, 7.97 and 2.29 with 9, 7 and 5 degrees of freedom. The null hypothesis in each case is that the corresponding s is correct and that the discrepancies between observed and expected number of counts occur by chance alone. On this basis, the probability of obtaining a poorer fit by chance is 0.2% for s=2, which indicates a very poor fit. For s=4 the probability of obtaining a poorer fit by chance is 33.3%, which is considered acceptable. Because of the high probability of a poorer fit for s=6 of 80.8%, it may be considered overfitted. The value of s is chosen to give a probability of a poorer fit roughly in the region of 30 to 60%, so as to provide a reasonably good fit using as few parameters as possible and at the same time avoiding the fitting of chance fluctuations in the data.

Figure 1 shows the plotted comparisons between the observed counts and fitted probability density functions for each of the three different degrees of the polynomial. In each case the smooth line is fitted while the heavy dots are observed values. The appearance of the curve for s=6 is consistent with its being

FIGURE 1

FITTED PROBABILITY DENSITY FUNCTIONS FOR S=2, S=4 AND S=6



overfitted. The curve for s=2 is the same as the normal probability density function with mean and variance estimated from the data.

Figure 2 shows the original data and the fitted curve for s=4 as in Figure 1 and also shows 95% confidence limits for the probability density function. Note that fitted values, upper and lower 95% confidence limits were only calculated at those values of x at which observations were available. The smooth curves were obtained by using a smoothing procedure based on cubic spline functions. Smooth curves could have also been obtained by using a higher level of discretization along the x-axis and then evaluating the fitted function and the limits at a larger number of x-values. Cubic splines were used for numerical convenience only and it is expected that both procedures would produce practically the same curve.

A second example related to uranium mill tailings was also used to illustrate application of the technique. Specifically, two data sets, one of 100 trials and one of 500 trials, of the variable DOSRN were provided by SENES Consultants Limited. The values of the variable DOSRN are estimates of the annual dose to an adult male living year round about 2 km downwind of the reference uranium tailings area described in SENES (1984) and which arises from radon gas released from the tailings area.

Tables 2 and 3 shown cell counts for the output variable DOSRN for 100 and 500 trials respectively. As shown the χ^2 goodness-of-fit test indicates that in both cases a model with s=2 is suitable. Figures 3 and 4 show the observed cell counts, the fitted probability density functions and the 95% confidence limits for these. As expected, the bands obtained at 500 trials are narrower than those at 100 trials.

FIGURE 2

95% CONFIDENCE LIMITS
FOR PROBABILITY DENSITY FUNCTION
CLASSROOM ECONOMIC STUDY (S=4)

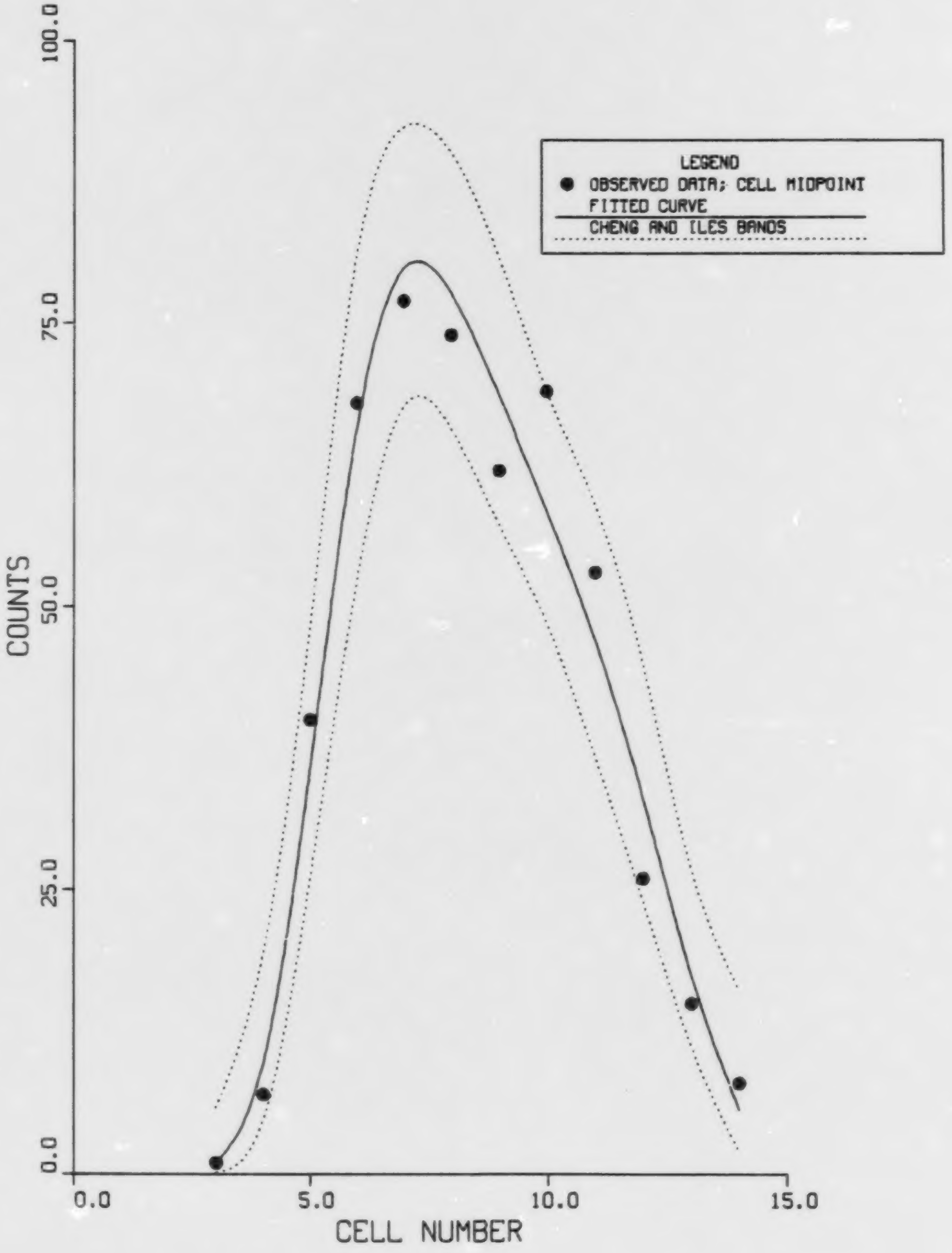


FIGURE 3

95% CONFIDENCE LIMITS
FOR PROBABILITY DENSITY FUNCTION
DOSRN 100 RUNS - 2 PARAMETER MODEL

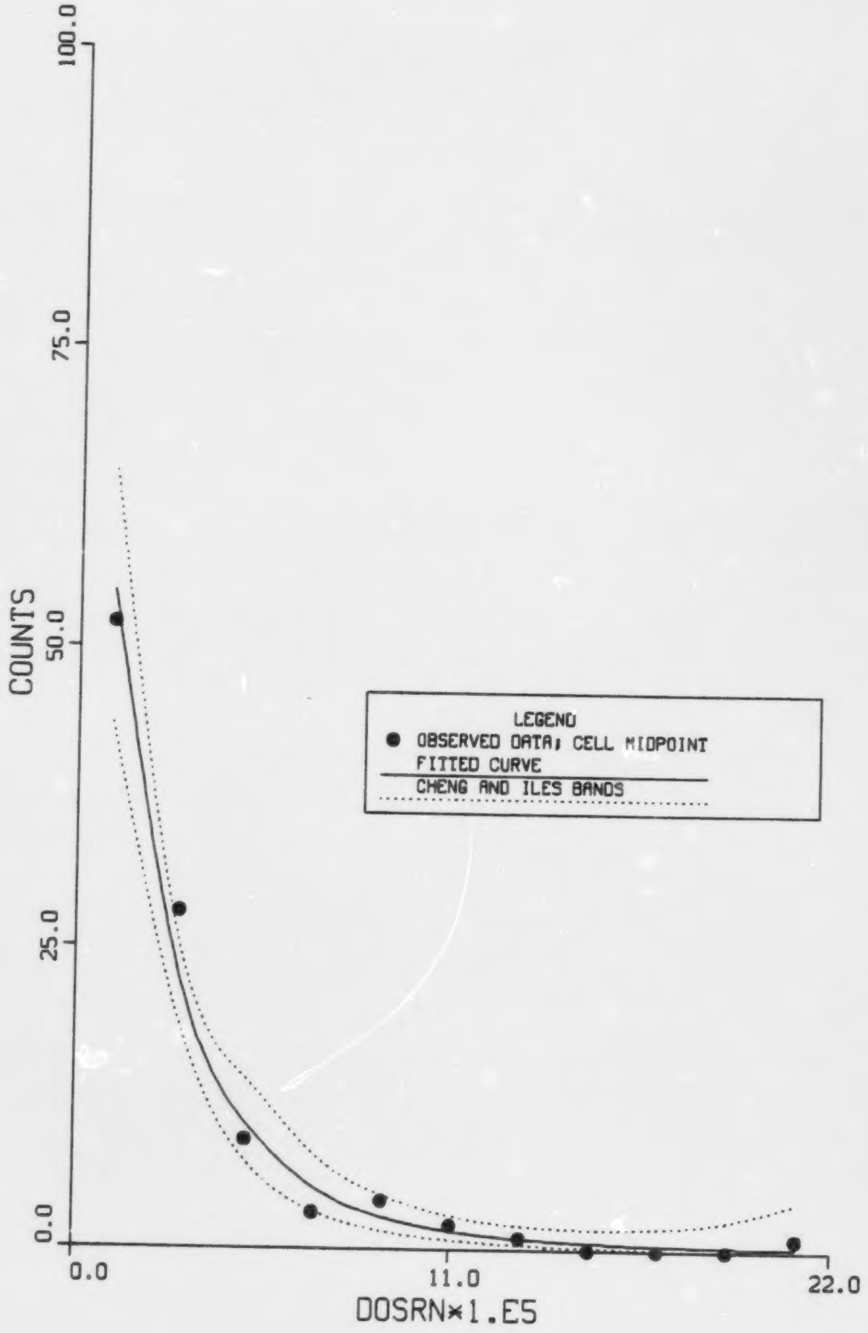
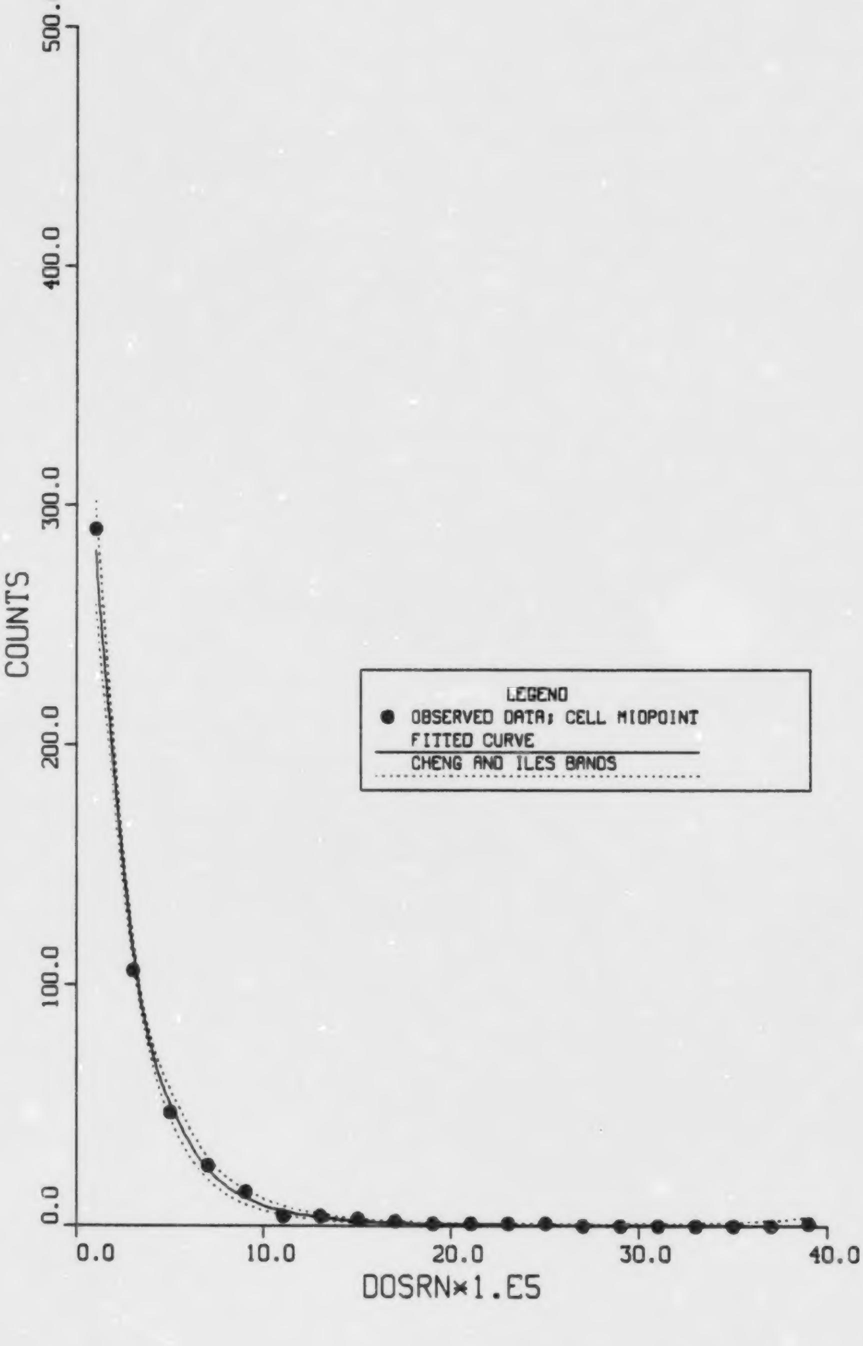


FIGURE 4

95% CONFIDENCE LIMITS FOR PROBABILITY DENSITY FUNCTION DOSRN 500 RUNS – 2 PARAMETER MODEL



Figures 5 and 6 show the corresponding cumulative distribution functions, including the 95% confidence bounds produced by this method as well as Kolmogorov bands. The procedure for obtaining the Kolmogorov bands is outlined in Appendix C.

It should be pointed out that the solid dots have a different meaning on the probability density function (pdf) plots (Figures 3 and 4), than they do on the cumulative distribution function (cdf) plots (Figure 5 and 6). On the pdf plots the dots represent observed data plotted at the cell midpoints. On the cdf plots they represent observed data plotted at the right hand cell endpoint. This arises from the fact that the ordinate is a measure of probability density on the one hand, while it measures cumulative probability on the other hand. In all of Figures 3 to 6, the cubic spline smoothing procedure described earlier was used.

7.0 CONCLUSIONS

A method has been developed for fitting a flexible generalized model to the output distribution from Monte Carlo type calculations.

A procedure has been developed for calculating confidence bands for probability density functions and cumulative distribution functions of Monte Carlo type output which is an extension of the approach formulated by Cheng and Iles and by Scheffé.

The model and confidence band procedure has been successfully applied to environmental assessment model output.

The bands for the cdf based on this method compare well with the Kolmogorov bands. They give narrower confidence bands which are however rigorously correct.

FIGURE 5

95% CONFIDENCE LIMITS
FOR CUMULATIVE DISTRIBUTION FUNCTION
DOSRN 100 RUNS - 2 PARAMETER MODEL

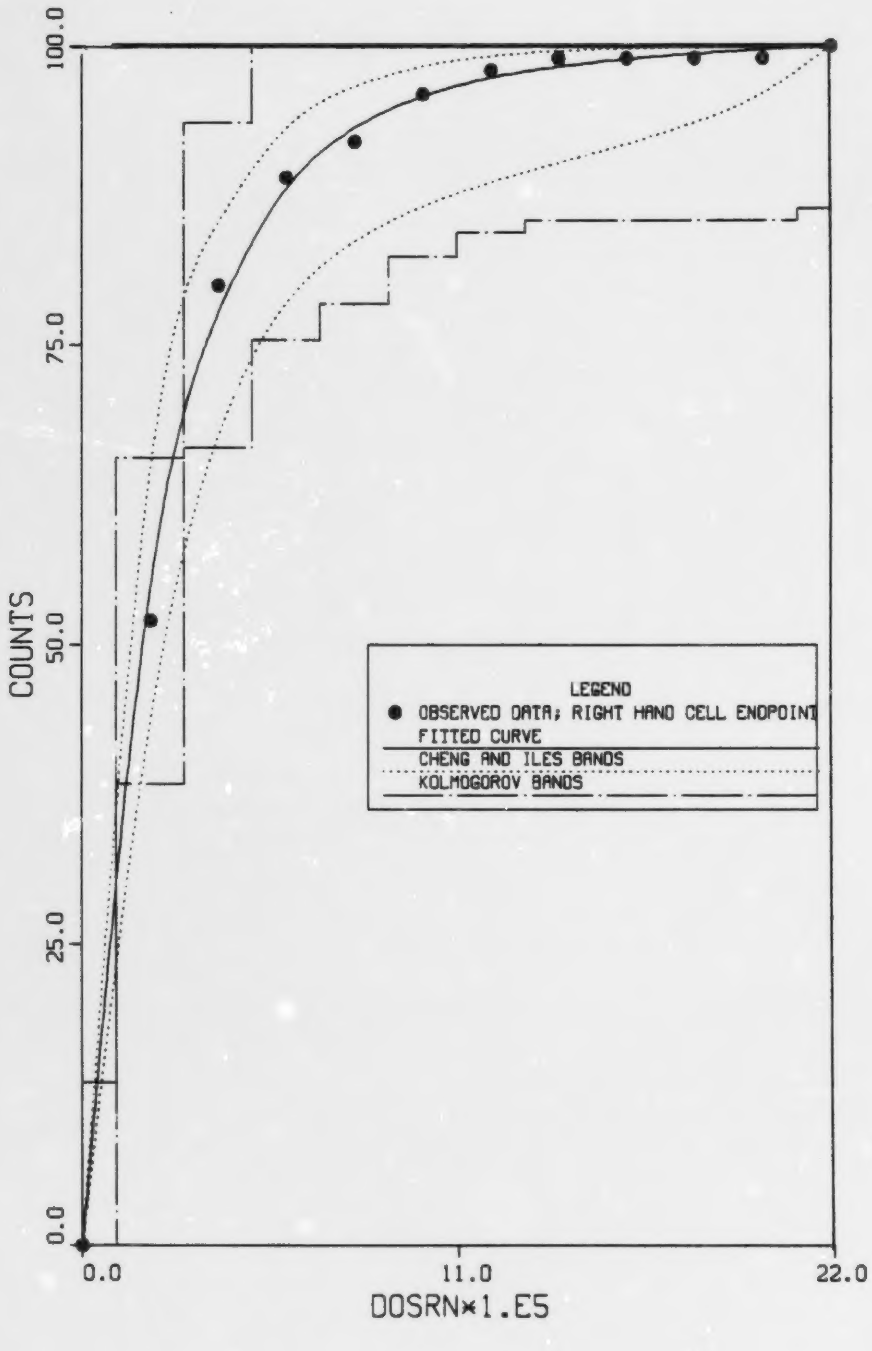
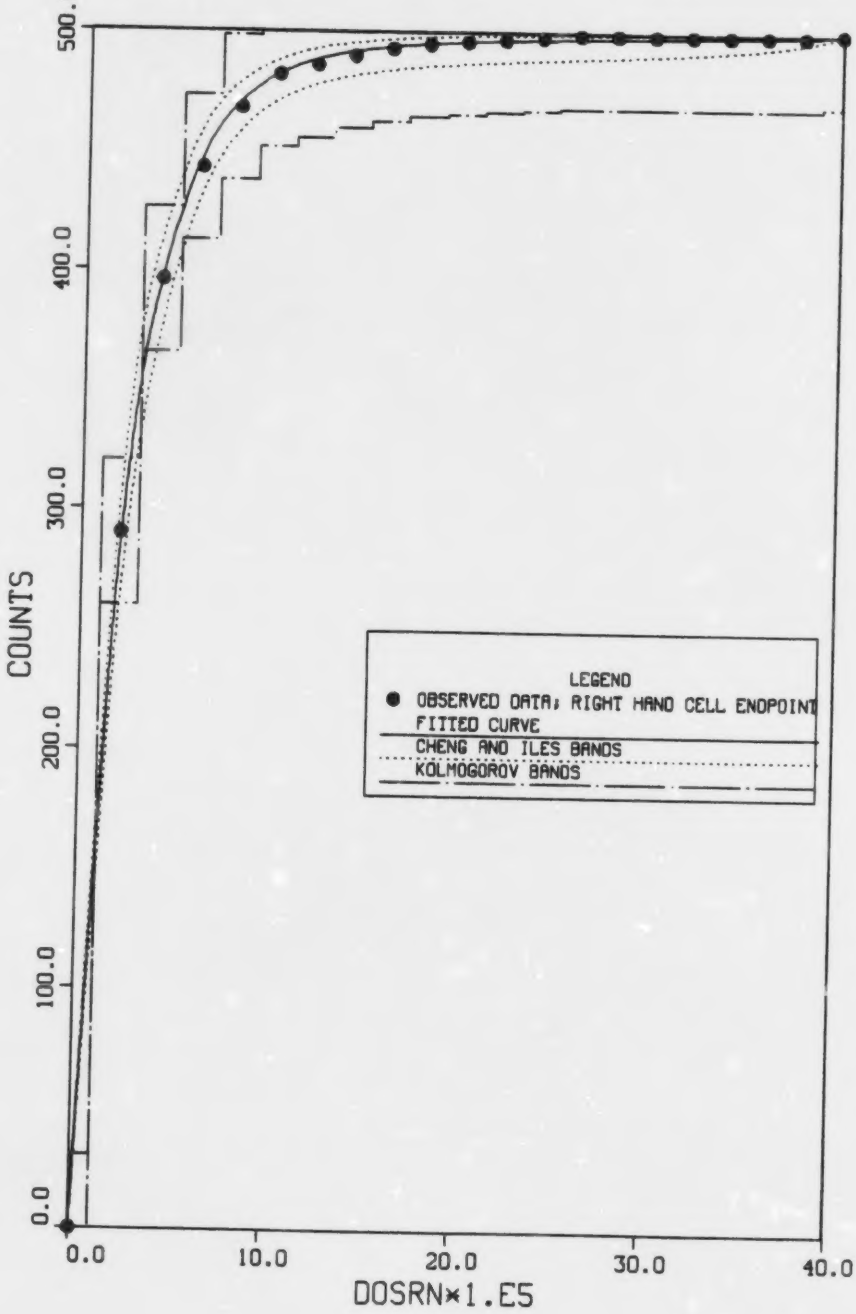


FIGURE 6

95% CONFIDENCE LIMITS FOR CUMULATIVE DISTRIBUTION FUNCTION DOSRN 500 RUNS - 2 PARAMETER MODEL



REFERENCES

Bard, Y., 1974. "Nonlinear Parameter Estimation." Academic Press.

Cheng, R.C.H., and Iles, T.C., 1983. "Confidence Bands for Cumulative Distribution Functions of Continuous Random Variables." Technometrics, Vol. 25, pp. 77-86.

Davies, O.L., and Goldsmith, P.L., 1977. "Statistical Methods in Research and Production." Longman Group Ltd., London.

Kendall, M., and Stuart, A., 1961. "The Advanced Theory of Statistics, Vol. 2: Inference and Relationship." 1st Edition, Hafner Publishing Co.

Kolmogorov, A.N., 1941. "Confidence Limits for an Unknown Distribution Function." Ann. Math. Statist., Vol. 12, pp. 461-463.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada - Phase I." Prepared for the National Uranium Tailings Program, Energy, Mines and Resources Canada, Ottawa, Ontario.

Scheffé, H., 1959. "The Analysis of Variance." John Wiley and Sons, New York.

Table 1

ILLUSTRATION OF CHENG AND ILES TECHNIQUE

Cell number	1	2	3	4	5	6	7	8	9	10	11	12
Actual counts (n_i)	1	7	40	68	77	74	62	69	53	26	15	8
Fitted (s=2)	8.2	17.9	33.6	53.3	72.1	82.8	80.9	67.2	47.4	28.5	14.5	6.3
Fitted (s=4)	1.0	10.0	35.8	65.7	79.9	78.0	69.2	59.0	47.7	33.8	18.1	5.9
Fitted (s=6)	.8	8.5	37.6	69.9	77.0	71.4	67.7	64.7	52.3	29.8	12.9	8.5
Upper 95% Confidence Boundary (s=4)	5.8	19.3	47.8	80.8	92.4	90.2	81.0	69.1	59.1	45.0	27.0	16.2
Lower 95% Confidence Boundary (s=4)	0.2	4.8	25.9	52.3	67.7	65.8	57.3	48.4	36.8	24.0	11.2	1.9

Table 2

VARIABLE DOSRN (100 TRIALS)

Cell Number	Observed Counts	Fitted (s=2)	Upper 95% Conf. Limits	Lower 95% Conf. Limits
1	52	54.7	64.6	43.6
2	28	23.0	25.7	19.1
3	9	10.4	14.1	7.2
4	3	5.1	7.8	3.1
5	4	2.7	4.5	1.5
6	2	1.5	2.8	0.8
7	1	0.9	2.0	0.4
8	0	0.6	1.7	0.2
9	0	0.4	1.9	0.1
10	0	0.3	2.5	0.0
11	1	0.3	4.0	0.0

Goodness-of-Fit Test

s	χ^2	d.f.*	Pr.**
1	3.607	3	30.7%
2	2.47	2	30.4%

* d.f. = degrees of freedom
** Pr. = probability of obtaining a poorer fit

Table 3

VARIABLE DOSRN (500 TRIALS)

Cell Number	Observed Counts	Fitted (s=2)	Upper 95% Conf. Limits	Lower 95% Conf. Limits
1	290	280.9	301.7	258.4
2	106	115.6	119.2	110.7
3	47	50.6	56.8	44.6
4	25	23.6	28.5	19.3
5	14	11.7	15.0	9.0
6	4	6.1	8.3	4.5
7	4	3.4	4.9	2.4
8	3	2.0	3.0	1.4
9	2	1.3	2.0	0.8
10	1	0.9	1.4	0.5
11	1	0.6	1.1	0.4
12	1	0.4	0.9	0.2
13	1	0.4	0.8	0.2
14	0	0.3	0.8	0.1
15	0	0.3	0.8	0.1
16	0	0.3	0.9	0.1
17	0	0.3	1.1	0.1
18	0	0.3	1.5	0.1
19	0	0.4	2.2	0.1
20	1	0.5	3.5	0.1

Goodness of Fit Test

s	χ^2	d.f.*	Pr.**
1	20.8	5	0.1%
2	3.24	4	53.0%

* d.f. = degrees of freedom
** Pr. = probability of obtaining a poorer fit

APPENDIX A: NEWTON-RAPHSON ITERATIVE SOLUTION

An iterative procedure for solving equation (10) by the Newton-Raphson method is as follows.

Define the (q+1)x1 vector function $\underline{F}(\underline{\Theta})$ as:

$$\underline{F}(\underline{\Theta}) = \sum_{i=1}^{m} x_i^{u-1} \left[\frac{n_i}{N} - p_i \right] \quad \text{uth element}$$

where all notation is as used previously and $p_i(x_i)$ is defined by equation (1). Define the (q+1)x(q+1) Jacobian matrix $\underline{J}(\underline{\Theta})$ as:

$$\underline{J}(\underline{\Theta}) = \frac{\delta \underline{F}}{\delta \underline{\Theta}} = \left[-\sum_{i=1}^{m} N p_i x_i^{u+v-2} \right] \quad \text{u,v element}$$

Then the iterative scheme consists of choosing a starting vector $\underline{\Theta}^{(1)}$ and calculating successive vectors from:

$$\underline{\Theta}^{(k+1)} = \underline{\Theta}^{(k)} - \underline{\delta}^{(k)}$$

where $\underline{\delta}^{(k)}$ is found by solving the following set of linear equations for $\underline{\delta}^{(k)}$

$$\underline{J}(\underline{\Theta}^{(k)})\ \underline{\delta}^{(k)} = \underline{F}(\underline{\Theta}^{(k)})$$

If q is large, say greater than 4, steps may be necessary to deal with the fact that $\underline{J}$ will likely be badly conditioned.

APPENDIX B: ANALYTICAL SOLUTION FOR r=1

Calculation of r̲ for s=1:

$$r = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} u \exp\{\Theta_o + \Theta_1 u\}\, du + u \right] \exp\{\Theta_o + \Theta_1 u\}\, du$$

$$= \int_{-\infty}^{\infty} u \exp\{\Theta_o + \Theta_1 u\}\, du \int_{-\infty}^{x} \exp\{\Theta_o + \Theta_1 u\}\, du + \int_{-\infty}^{x} u \exp\{\Theta_o + \Theta_1 u\}\, du$$

$$= \exp\{\Theta_o\} \left[\int_{-\infty}^{x} u \exp(\Theta_1 u)\, du \int_{-\infty}^{x} \exp(\Theta_1 u)\, du + \int_{\infty}^{x} u \exp(\Theta_1, u)\, du \right]$$

For s=1, problems arise when trying to integrate to $-\infty$. Therefore use the smallest value of x say +a.

$$\exp(\Theta_o) = \frac{1}{\int_{+a}^{\infty} \exp(\Theta_1 u)\, du} = -\Theta_1 \exp(-a\Theta_1)$$

$$r = -\frac{1}{\Theta} \exp(-a\Theta_1) \left[\left(\frac{\exp(\Theta_1 x + a\Theta_1) - \exp(2a\Theta_1)}{\Theta_1} \right) (1-a_{\Theta_1}) + \exp(\theta_1 x)(\theta_1 x - 1) + \exp(a\theta_1)(1-a\theta_1) \right]$$

APPENDIX C: KOLMOGOROV CONFIDENCE BANDS

For 95% confidence bands and for a number of trials larger than 50,

$$d_{.95} = 1.3581\ n^{-1/2};\quad n = \text{number of trials}$$

For DOSRN 100: $d_{.95} = .13581$
For DOSRN 500: $d_{.95} = .06076$

To plot the bands on graphs, multiply d_α by the sample size.

DOSRN 100: $d_{.95} = 13.6$
DOSRN 500: $d_{.95} = 30.4$

Kolmogorov bands are then obtained by plotting a band $\pm\ d_{.95}$ about the observed cumulative frequency function.

DISCUSSION PAPER ON

SENSITIVITY/UNCERTAINTY ANALYSIS AND EVALUATION OF TIME RESPONSE OF A MODEL OF RADIONUCLIDE ACCUMULATION IN SOIL

by
D.I. Beals, D.B. Chambers and B.E. Halbert
SENES Consultants Limited
June 1985

Abstract

A multi-component model describing radionuclide accumulation in a soil profile from atmospheric deposition is developed and used to investigate methods for treating time-varying input parameters and to demonstrate the usefulness of sensitivity and uncertainty analysis techniques. Model response is examined in a series of probabilistic simulations in which uncertain, non-time-variant parameters are sampled randomly at the beginning of each trial and time-variant parameters are either:

i) sampled at the beginning of each time step from a distribution of annual means (applicable to a one year time step)
ii) sampled at the beginning of each timestep from a distribution of ten year means (applicable to a ten year time step),
iii) sampled at the beginning of each trial from a distribution of annual means (that is, sampled in the same manner as the non-time-variant parameters)
iv) held constant at their mean annual values for all trials (that is, treated as constants).

All methods yielded similar predicted mean soil concentrations. Other statistical attributes were found to vary considerably between the approaches.

Examination of several sensitivity and uncertainty analysis techniques, ranging from quite simple measures such as sensitivity and uncertainty indices to the sophisticated regression analysis method, yielded similar conclusions. Because the applicability of the sensitivity and uncertainty analysis techniques is dependent on model structure, a flexible creative approach using all available tools is recommended.

Table of Contents

Page

1.0 INTRODUCTION 1

2.0 SOIL LEACHING MODEL 2
  2.1 Soil Model Development 2
  2.2 Probabilistic Implementation of the Soil Model 7

3.0 TREATMENT OF TIME AND TIME VARIANT PARAMETERS 8
  3.1 Types of Input Parameters 9
  3.2 Some Different Ways of Treating Time Variant Parameters 10
  3.3 Some Examples 12
  3.4 Suggestions for Stepping Through Time 13

4.0 SOME TECHNIQUES FOR SENSITIVITY AND UNCERTAINTY ANALYSIS 15
  4.1 Sensitivity Index 15
  4.2 Graphical Presentation 16
  4.3 Uncertainty Index 16
  4.4 Regression Analysis Techniques 18

5.0 SENSITIVITY AND UNCERTAINTY ANALYSIS OF THE SOIL MODEL 22
  5.1 Sensitivity Index and Graphical Technique 22
  5.2 Uncertainty Index 24
  5.3 Regression Methods 24
  5.4 Discussion of Results 27

6.0 CONCLUSIONS 28

REFERENCES 31

APPENDIX A - SOME NOTES ON REGRESSION ANALYSIS TECHNIQUES 33

List of Tables

		Follows Page
2.1	Input Parameter Distributions and Statistical Descriptors	7
3.1	Summary Statistics of Soil Layers 1 and 2 Nuclide Concentration for 4 Alternative Approaches to Stepping Through Time	12
3.2	Summary Statistics for Radionuclide Concentrations Soil Layer 2 at 100 Years for 2 Approaches to Stepping Through Time	13
5.1	Summary of Various Sensitivity and Uncertainty Indices for Soil Model Inputs	22
5.2	Ranks of Various Sensitivity and Uncertainty Indices for Soil Model Inputs	22

List of Figures

Follows Page

2.1 Schematic Representation of Soil Deposition-Leaching Concepts 2

3.1 Comparison of Three Methods of Treatment of Time For Soil Layer 1 For 100 Years 12

3.2 Comparison of Three Methods of Treatment of Time For Soil Layer 2 for 100 Years 12

5.1 Concentration and Time to Steady State Precipitation (Soil Layer 1) 22

5.2 Concentration and Time to Steady State vs. Kd (Soil Layer 1) 22

5.3 Concentration and Time to Steady State vs. Dry Bulk Density (Soil Layer 1) 22

5.4 Graphical Description of Sensitivity Index for Parameter X_1 23

1.0 INTRODUCTION

In the Phase I study on the application of probabilistic assessment techniques to uranium mine facilities (SENES 1984), a set of mathematical models was developed to describe six selected exposure pathways by which radioactivity could be released from a tailings area and result in human exposure. These six pathways were then used in a preliminary probabilistic assessment of a reference uranium mill tailings site.

During the Phase I study, various techniques were used to investigate the question "which parameters/pathways have the greatest effects on model predictions". However, detailed examination of the various statistical analysis and regression techniques in use for sensitivity/ uncertainty analysis was outside the scope of the Phase I study. One objective of this discussion paper, which forms part of the Phase II studies, is to examine, in a tutorial form, some of the tools that can be used to carry out sensitivity/ uncertainty analyses of a probabilistic model.

Another aspect which was not considered in Phase I but is of potential interest, is the way in which airborne radioactivity deposits on the ground surface and accumulates in the upper soil horizons. Often this process is modelled simply as the balance between deposition from the air and environmental removal (typically treated as an arbitrary lumped parameter). In this paper, the mechanisms controlling the movement of air-deposited radioactivity in the upper soil horizon are modelled explicitly, namely the downward migration of the deposited radioactivity arising from leaching and radionuclide sorption on soil particles. The soil leaching model is subsequently used to illustrate the application of some of the techniques, including regression analysis, available for sensitivity/uncertainty analysis.

One further aspect of interest to the Phase II studies is also examined, namely, how should propagation through time be incorporated in a probabilistic assessment. The soil model is used to illustrate how time-dependent behaviour, and in particular how parameters which vary with time can be treated.

2.0 SOIL LEACHING MODEL

This section outlines the development of a model that describes the movement of radionuclides deposited as particulate matter from the air to the surface of the soil. The model includes atmospheric deposition, leaching and downward transport of radionuclides into the soil profile.

2.1 Soil Model Development

Figure 2-1 indicates the major factors at work in balancing radioactivity in a surface soil layer. Three factors are considered in the soil model developed in this section:

1) incoming airborne radioactivity;

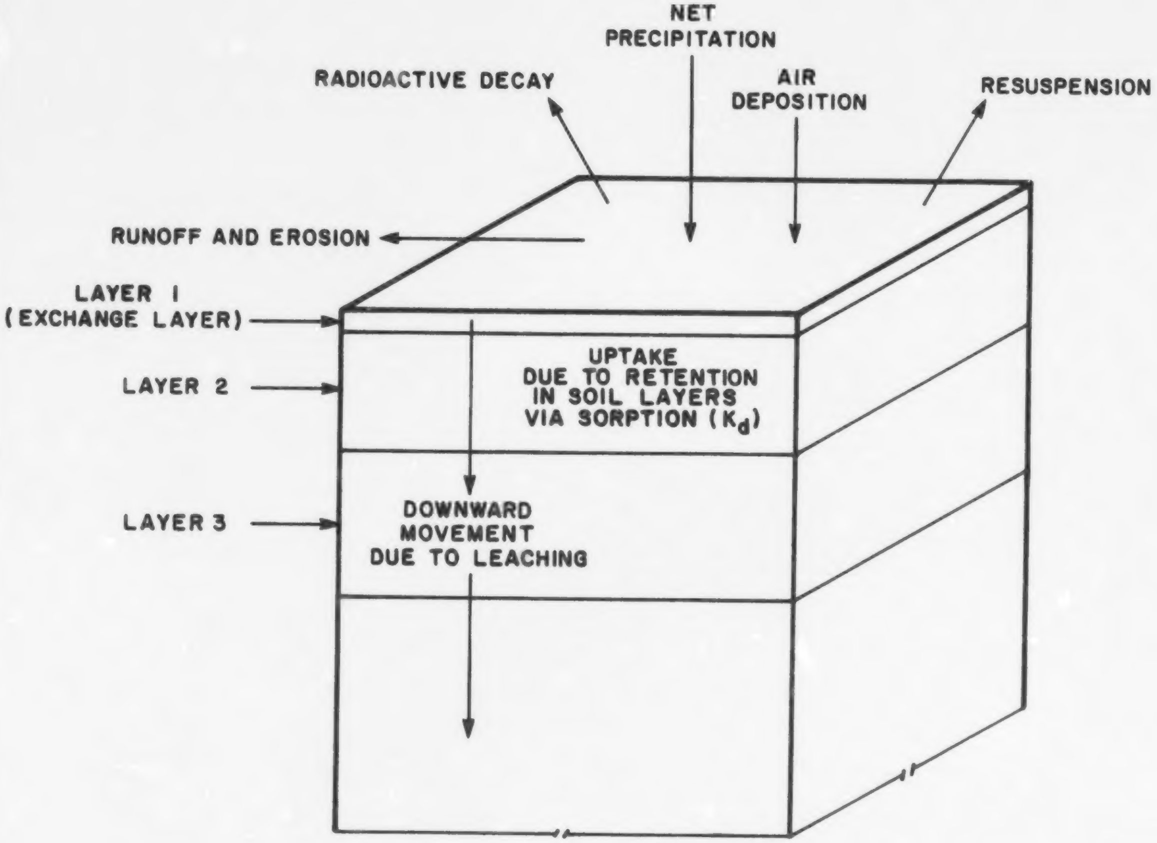
2) radioactive decay; and,

3) radioactivity leaving the soil compartment in solution in the soil water via leaching.

Surface runoff, resuspension and erosion are not accounted for explicitly in the model. Since the quantity of interest is the radionuclide concentration in the upper soil horizons these assumptions are conservative (i.e. result in high estimates of

FIGURE 2.1

SCHEMATIC REPRESENTATION OF SOIL DEPOSITION - LEACHING CONCEPTS



residual radionuclide levels). The activity balance in a surface soil layer receiving radioactivity deposited from the air can be mathematically formulated as:

$$dZ_1 = C_A V_g Adt - C_w PAdt - \lambda_R Z_1 dt \qquad (2\text{-}1)$$

where

Z_1 = radionuclide content of soil layer 1 (Bq)

A = soil area (m^2) [taken as 1 m^2 in model calculations]

C_A = radionuclide concentration in air ($Bq.m^{-3}$)

V_g = particulate settling velocity ($m.yr^{-1}$)

t = time (yr) (1 year time increments unless otherwise specified))

C_w = radionuclide concentration in soil water ($Bq.m^{-3}$) (water is assumed to be the mobile phase)

P = total precipitation less evapotranspiration ($m.yr^{-1}$) (i.e. net precipitation)

λ_R = radioactive decay constant (yr^{-1})

The radionuclide content of the surface soil layer is the sum of the activities in the solid, liquid and gaseous phases of the soil or:

$$Z_1 = C_w V_w + C_s\ \rho_s AD + C_g V_a \qquad (2\text{-}2)$$

where

V_w = volume of water in pore space (m^3)

ρ_s = soil dry bulk density ($kg.m^{-3}$)

C_s = radionuclide concentration in solid phase ($Bq.kg^{-1}$)

V_a = volume of air in soil(m^3)

C_g = radionuclide concentration in gaseous phase ($Bq.m^{-3}$)

A = soil area(m^2)

D = soil depth(m)

The activity in the gaseous phase for non volatile nuclides is assumed to be zero.

The distribution of radionuclides between the solid and liquid phases is assumed to be in equilibrium and, thus, characterized by the distribution coefficient $K_d (m^3.kg^{-1})$:

$$K_d = C_s/C_w \qquad (2\text{-}3)$$

The volume of water in the soil pore space (V_w in m^3) can be calculated as a function of the water content of the soil (W in kg water/kg soil) corrected for the relative densities of the solid and liquid fractions:

$$V_w = WAD\rho_s/\rho_w \qquad (2\text{-}4)$$

where ρ_w=density of water = 998.2 $kg.m^{-3}$ at 20°C.

Combining equations (2-2),(2-3) and (2-4), the radionuclide content of the surface soil layer for non-volatile nuclides is given by:

$$Z_1 = C_w \, WAD \, \rho_s/\rho_w + K_d \, C_w \, \rho_s AD$$

and solving for C_w we get:

$$C_w = \frac{Z_1}{\rho_s \; AD(W/\rho_w + K_d)} \qquad (2\text{-}5)$$

Substituting equation (2-5) into equation (2-1):

$$dZ_1 = C_A V_g Adt - \left[\frac{Z_1 Pdt}{\rho_s D(W/\rho_w + K_d)}\right] - \lambda_R Z_1 dt$$

$$dZ_1 = (\beta_o - \beta_1 Z_1 - \lambda_R Z_1)dt \qquad (2\text{-}6)$$

where

$$\beta_o = C_A V_g A$$

$$\beta_1 = \frac{P}{\rho_s \; D \; (W/\rho_w + K_d)}$$

Equation (2-6) has the solution:

$$Z_1(t) = Z_1(o)\exp[-(\beta_1 + \lambda_R)t] + \frac{\beta_o}{\beta_1 + \lambda_R}[1 - \exp[-(\beta_1 + \lambda_R)t]] \qquad (2\text{-}7)$$

where the activity in the layer is $Z_1(o)$ at $t=0$.

The water concentration (C_w) is obtained from equation (2-5) after solving for $Z_1(t)$ in equation (2-7). From either equation (2-6) or (2-7) the steady-state solution (Z_1^{ss}) for the uppermost soil layer is:

$$Z_1^{ss} = \frac{\beta_o}{\beta_1 + \lambda_R} \qquad (2\text{-}8)$$

Equation (2-7) was used to calculate the concentration in the 1st layer. Successive soil layers were dealt with in the manner described below.

The activity leaving the first layer is the time interval ranging from t to t+Δt, ($\Delta Z_1(\Delta t)$) may be calculated from:

$$\Delta Z_1(\Delta t) = [(Z_1(t) + C_A V_g A \Delta t) - Z_1(t+\Delta t)]/\Delta t \qquad (2\text{-}9)$$

where

Δt = time increment used in model (yr).

$Z_1(t)$ = amount of activity in layer 1 at beginning of time increment (calculated from equation (2-7) for previous time increment) (Bq).

$Z_1(t+\Delta t)$= amount of activity in layer 1 at end of time increment (calculated from equation (2-7)) (Bq)

$C_A V_g A\ \Delta t$= the amount of the radionuclide deposited from the air to soil layer 1 in time Δt (Bq).

Since the numerator of equation (2-9) is the activity leaving in the time Δt it must be divided by Δt to get the activity leaving in one year.

The parameter ΔZ_1 would then be the radionuclide input (during Δt) to the 2nd layer. The initial inventory of the radionuclide present in soil layer 2 at the start of the time step, namely $Z_2(t)$, would be the radionuclide inventory left over from the previous time step (note that in the present formulation at $t=0$; $Z_2(0)=0$). Thus, the amount leaving the 2nd soil layer as input to the 3rd soil layer ($\Delta Z_2(\Delta t)$) per unit time would be:

$$\Delta Z_2\ (\Delta t) = [Z_2(t) + \Delta Z_1(\Delta t) - Z_2(t+\Delta t)]/\Delta t \qquad (2\text{-}10)$$

Calculations for the 3rd layer (and any successive layers) would be done in the same manner. A computer program was written using the approach outlined above to carry out the simulations and analyses discussed in the following sections.

Since other loss mechanisms, such as runoff, erosion and resuspension are neglected, the activity remaining in the soil column may be overestimated. On the other hand, deposition of eroded material from adjacent parcels of land may offset the potential losses noted above. In the model structure presented above, the radioactivity level in each soil layer increases to some steady level where input balances output.

2.2 Probabilistic Implementation of the Soil Model

The soil model was run in a probabilistic fashion using simple random sampling. The distributions assigned to the input parameters, including the distribution types and statistical descriptors, are presented in Table 2.1. The reference simulation was based on radium for a 100 year simulation period using 1 year time steps and 100 Monte Carlo trials for each simulation.

Prior to carrying out the probabilistic analysis, deterministic simulations were used to investigate the behaviour of the model. A 1 cm upper layer of organic material was used to define an atmospheric exchange layer. The deterministic trials indicated that steady state was easily achieved within the 100 year simulation period. The upper soil layer is the focus of the sensitivity/uncertainty analysis discussed in Section 4.0.

Layer 2 was defined to have a 15 cm depth and to have a higher bulk density than layer 1. The technical approach to sensitivity and uncertainty analysis discussed in Section 4.0 would be similar for layer 2.

3.0 TREATMENT OF TIME AND TIME VARIANT PARAMETERS

This section discusses some considerations regarding the way in which the soil system, described by the soil model detailed in Section 2, is anticipated to evolve with time. The role of uncertain parameters which vary over time is contrasted to the role of parameters which are uncertain but do not vary with time. The effect of different length time steps (Δt's) is also

Table 2.1

Input Parameter Distributions and Statistical Descriptors

Symbols	Parameter (Units)	Distribution	Nominal Value	Minimum	Maximum
-	Precipitation* ($m.yr^{-1}$) varies annually	normal	0.9 (= 0.14)	0.5	1.2
-	Evapotranspiration ($m.yr^{-1}$)	constant	0.4	-	-
V_g	Deposition Velocity ($m.yr^{-1}$)	triangular	$4.1x10^5$	$6.3x10^4$	$1.4x10^6$
λ_R	Decay constant radium (yr^{-1})	constant	$4.3x10^{-4}$	-	-
C_A	Concentration of Radium in Air ($Bq.m^{-3}$) (Layer 1 and 2)	triangular	$1.5x10^{-3}$	$1.5x10^{-5}$	$1.5x10^{-2}$
ρ_s	Soil Bulk Density $Kg.m^{-3}$ - Layer 1	uniform	235	20	450
	- Layer 2	triangular	$1.4x10^3$	$1.2x10^3$	$1.8x10^3$
W	Soil Water Content - Layer 1	uniform	1.075	0.75	1.4
	- Layer 2	uniform	0.21	0.05	0.36
K_d	Distribution Coefficient Layer 1 and 2 $m^3.Kg^{-1}$	triangular	0.30	0.01	0.50

* Net Precipitation (P) = precipitation-evapotranspiration

discussed.

First, let us consider how the time dependent behaviour of the soil system is calculated. In a deterministic evaluation, suitable parameter values are first entered into the model and then the value of the time dependent variable(s) is calculated as the system evolves in time from a specified initial state. This approach may also be taken in a probabilistic evaluation of the model. In this case however, the actual parameter values are not uniquely specified, but rather for each trial a set of parameter values are selected by randomly sampling from a suitable distribution of possible values for each parameter. The system is then allowed to evolve through time with the parameter values fixed at the initially selected values. In the next trial, new parameter values are selected by randomly sampling again from the specified parameter distributions, and the system is allowed to evolve once more in time. The process is repeated until the desired number of trials have been carried out.

In essence this procedure determines a family of (time) curves, one for each trial. Each curve represents the behaviour of the time dependent variable with respect to the particular set of inputs randomly selected for independent parameters.

Consider now a particular time dependent variable such as the concentration of radionuclide in the uppermost soil layer (i.e. layer 1). Each trial will produce a different curve of concentration versus time. At any specific time the actual values calculated for the soil concentration may differ widely from one trial to the next. Values at a particular time from all of the trials can be used to construct a frequency diagram and calculate various sample statistics such as the mean and variance. The "sample" in this instance is of course the ensemble of individual concentrations obtained from the same time interval for each trial.

Unfortunately, the situation is not quite as simple as indicated in the foregoing discussion, since some of the input parameters are not only uncertain, but vary from year to year. Variability from year to year in input parameters gives rise to uncertainty of a nature no different in principal from that which exists when the input parameters are uncertain but non-time-varying. In the latter case, the uncertainty arises from a lack of knowledge of a constant (or from spatial variability) in the other from lack of ability to predict a time-variant parameter. Uncertainty is present in both cases and is communicated by the output distribution. In both cases they describe how strongly one is entitled to believe in a particular consequence (soil radionuclide level in this instance) in a particular year. In one case, however, the consequence will remain constant from year to year, in the other it will vary. The possible ramifications of this latter aspect are discussed below.

3.1 Types of Input Parameters

Three distinct types of input parameters are used in the soil model when it is run in the probabilistic mode:

1. Physical constants: These parameters are input as fixed values and are held constant at the same value during all trials. An example of this type of parameter is the radioactive half-lives for the nuclides.

2. Uncertain but non-time-variant parameters: These input parameters are selected from a specified distribution each trial and remain constant for a given trial. Soil bulk density is an example of such a parameter.

3. Uncertain time-variant parameters: These parameters are not only uncertain but vary from year to year and thus change for each time step. Precipitation varies from year to year and is an example of a time-varying parameter.

3.2 Some Different Ways of Treating Time-Variant Parameters

Uncertain time-varying parameters can be treated in several different ways including:

1. Uncertain time-varying parameters could be sampled from a suitable distribution for each time increment. Annual precipitation for example, is assumed in this analysis to be normally distributed with a mean of 0.9 $m.yr^{-1}$ and a standard deviation of 0.14 (see Table 2.1). If a time increment (Δt) of one year is used in the simulation then a value for mean annual precipitation could be sampled directly from this distribution for each time increment.

2. If, however, a larger time increment is used, the distribution of mean annual precipitation requires adjustment. The approach taken in the this report is based on consideration of the sampling distribution of the mean. Specifically, consider mean annual precipitation to be a random variable with a population mean of μ_p and a variance σ_p^2, and that $\bar{P}$ is the sample mean from a random selection of n values of P and S_p^2 is the sample variance. These sample statistics have the following properties:

$$E(\bar{P}) = \mu_p$$
$$Var(\bar{P}) = \sigma_p^2/n$$
$$E(S_p^2) = \sigma_p^2$$

Consider for example, stepping through time in 10 year increments. The sampling distribution of mean annual precipitation would, for practical purposes, be normal with a mean μ_p and a variance $\sigma_p{}^2/n$. In terms of our original sample of values of P, the sampling distribution of the mean is specified by the mean $\bar{P}$ and the standard error of the mean: SE $(\bar{P}) = S_P/\sqrt{n}$.

3) The foregoing discussion also suggests that precipitation can be treated as a constant, set at its (long term) mean annual value

4) Precipitation can be treated (and commonly is) as an uncertain but non-time-variant parameter and is selected from its annual distribution at the beginning of each trial. This value would prevail for the entire trial. There is at least surficially some initial attraction to the modeller to adopt this method since precipitation is inherently uncertain. If annual precipitation is assumed to be statistically independent from year to year and a single precipitation value with a probability of 0.1, is selected at the beginning of a 100 year modelling period, the probability of 100 consecutive 0.1 probability precipitation years is $(0.1)^{100}$. Clearly there is only a vanishingly small chance of such an event. The modeller needs to be concerned with whether or not such a situation would lead to unrealistic maxima or minima, particularly when dealing with a sensitive uncertain parameter.

3.3 Some Examples

The soil model was used to illustrate each of the four approaches discussed above. The model was run probabilistically with 500 trials being realized for each approach. A summary of the statistical descriptors for the calculated concentrations in soil layers 1 and 2 in the 100th year of the simulation is given in

Table 3.1 The similarity of the four techniques in terms of all statistical descriptors is pronounced. Only the minimum and maximum values differ noticeably among the approaches.

Examples of the time behaviour of the first three cases discussed in the previous section are illustrated in Figure 3.1 for soil layer 1 and Figure 3.2 for soil layer 2. The relatively extreme transient behaviour of layer 1 in the annually varying case is noteworthy, particularly when compared to the behaviour of layer 2 for the same situation. Layer 1 is thin (0.01 m), has a low bulk density and high organic composition, all of which contributes to its low "damping" capacity. Conversely layer 2 which is 0.15 m thick, has a high bulk density and demonstrates a high damping capacity. Note the less radical transient behaviour of the 10 year time step curve in Figure 3.1. As seen on Figure 3.2, layer 2's behaviour is virtually identical in all three cases. The curve for the ten year time step example in Figure 3.1 is above the mean value curve due to random error in sampling the ten mean annual precipitation values and is not significant.

It was initially hypothesized that the output distribution would be distorted and minimum and maximum values would be more extreme when precipitation was treated as a non-time-variant but uncertain parameter (Case 4). Since there was little numeric difference between the time-variant and non-time-variant cases studied (i.e. cases 1 and 4) it was decided to test the two methods further by inclusion of both air concentration (C^A) and precipitation as uncertain time-variant and non-time-variant parameters. It was conjectured that because precipitation was not a highly variable parameter the effect it had was not as significant as would be observed with a more highly variable parameter such as the air concentration (see Table 2.1 for the parameter's statistical descriptors).

Table 3.1

Summary Statistics of the Soil Layer Nuclide Concentration for 4 Alternative Approaches to Stepping Through Time*

		Mean (Bq/m^3)	Standard Deviation	Geometric Mean (Bq/m^3)	Geometric Standard Deviation	Minimum (Bq/m^3)	Maximum (Bq/m^3)
Time Varying Precipitation, Annual Increments (Case 1)	Layer 1	4.6×10^5	5.6×10^5	2.4×10^5	3.6	2.9×10^3	4.8×10^6
	Layer 2	1.4×10^6	1.2×10^6	9.8×10^5	2.5	5.2×10^4	6.8×10^6
Time Varying Precipitation, 10 Year Increments (Case 2)	Layer 1	4.2×10^5	4.7×10^5	2.2×10^5	3.6	2.0×10^3	3.1×10^6
	Layer 2	1.3×10^6	1.1×10^6	9.1×10^5	2.6	2.3×10^4	5.6×10^6
Constant Precipitation (Mean Annual) (Case 3)	Layer 1	4.4×10^5	5.0×10^5	2.3×10^5	3.6	1.3×10^3	3.0×10^6
	Layer 2	1.4×10^6	1.1×10^6	9.8×10^5	2.5	6.2×10^4	6.6×10^6
Non-Time Variant But Uncertain Precipitation (Case 4)	Layer 1	4.3×10^5	5.3×10^5	2.2×10^5	3.6	2.6×10^3	4.7×10^6
	Layer 2	1.3×10^6	1.1×10^6	9.4×10^5	2.5	4.8×10^4	6.5×10^6

* statistics for 100th year.

FIGURE 3·1

COMPARISON OF THREE METHODS OF TREATMENT OF TIME FOR SOIL LAYER I FOR 100 YEARS

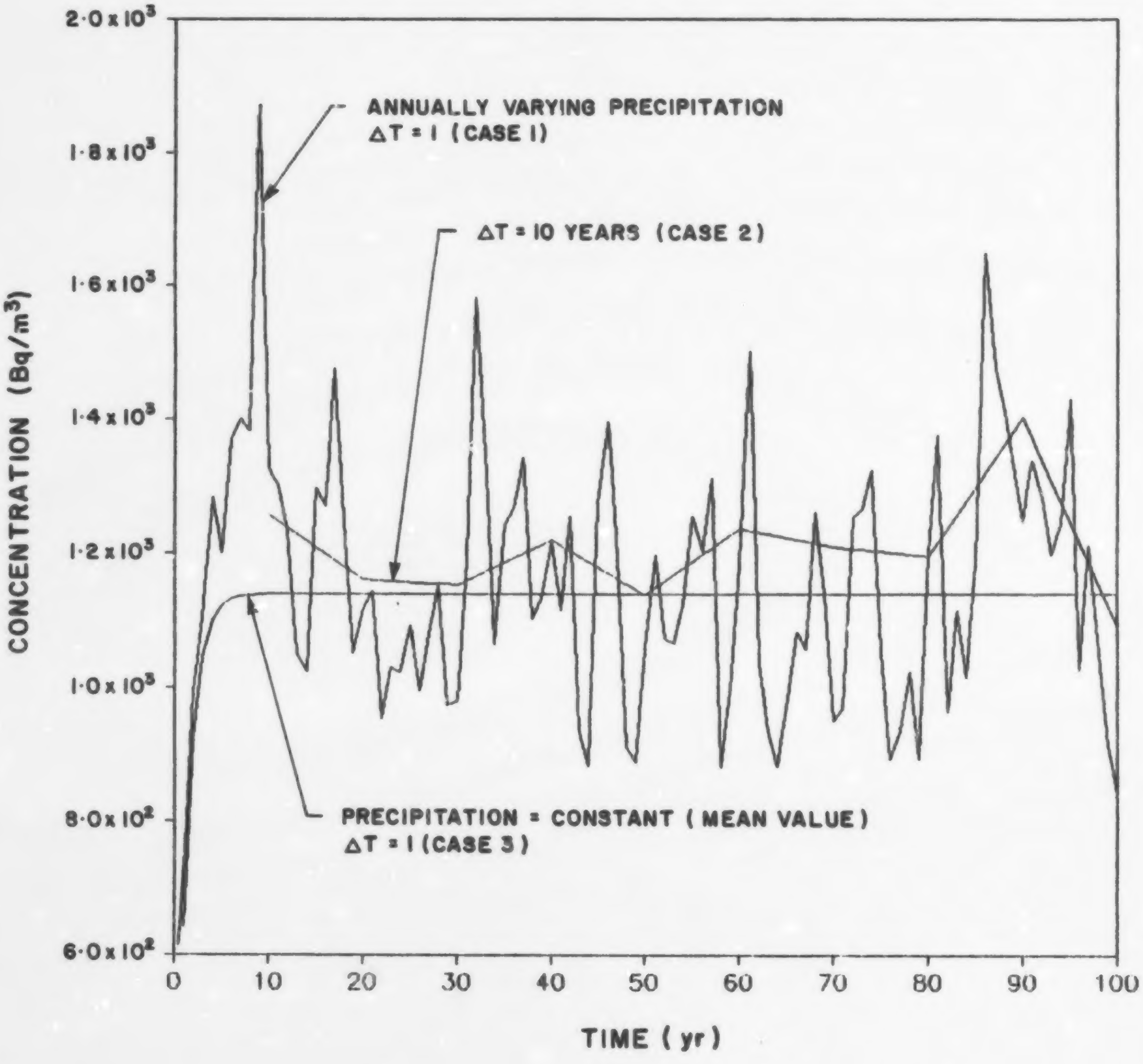
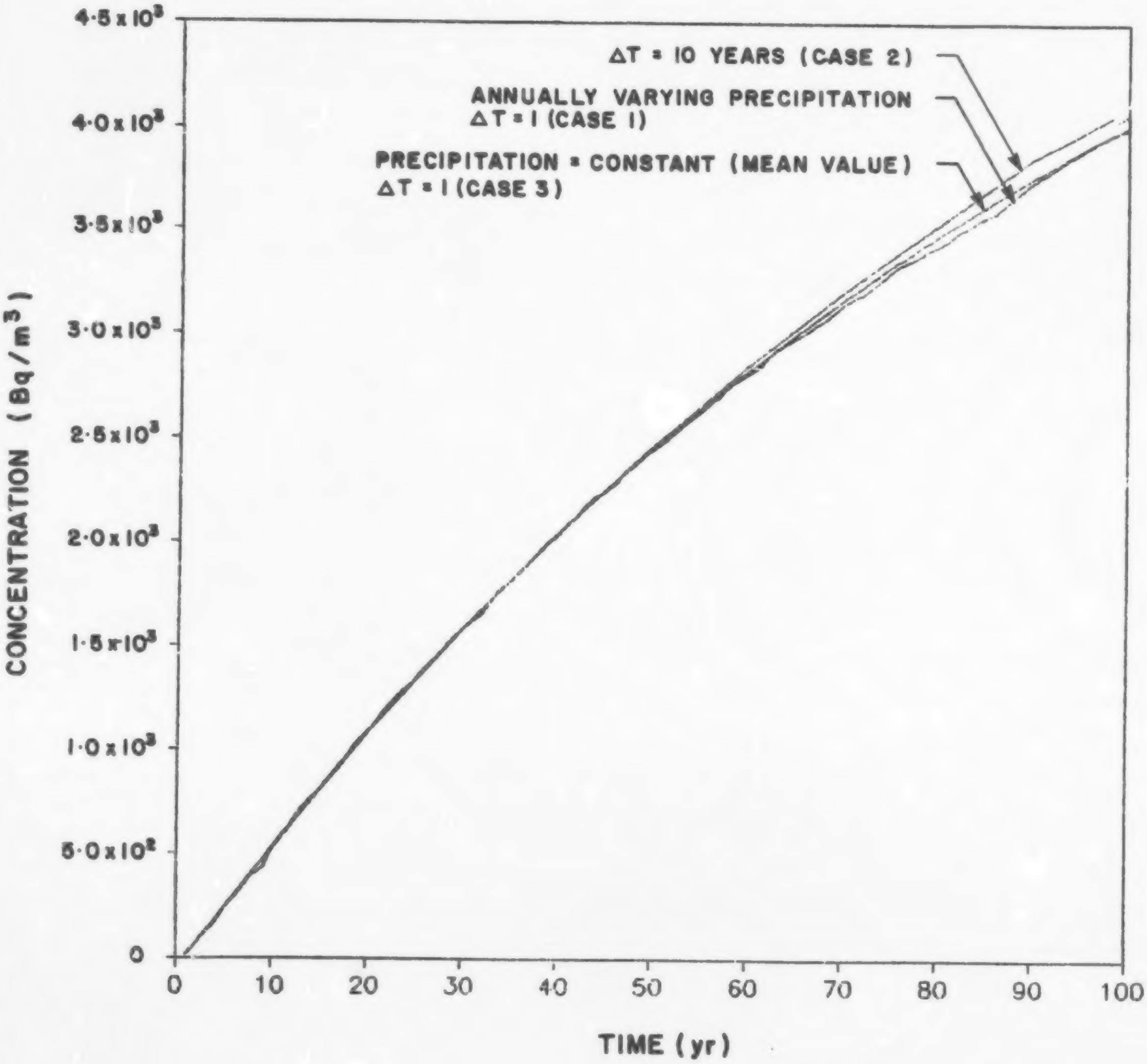


FIGURE 3·2

COMPARISON OF
THREE METHODS OF TREATMENT OF TIME
FOR SOIL LAYER 2 FOR 100 YEARS



The statistical descriptors for radionuclide concentration in soil layer 2 are shown in Table 3.2. It should be noted that while the mean values are similar, all the other descriptors differ more markedly than for the situation shown in Table 3.1. Indeed the higher standard deviation and geometric standard deviation and the more extreme values (i.e. the minimum and maximum values) calculated for the uncertain non-time-variant case indicate a substantially more dispersed population. This lends creedence to the hypothesis stated earlier that treatment of uncertain time-variant parameters as uncertain non-time variant parameters may distort the output distribution and, in particular, it has potential to produce unrealistic minimums and maximums.

3.4 Suggestions for Stepping Through Time

There are several considerations that the modeller must weigh and balance when considering how time and parameters which vary in time, should be incorporated into the probabilistic modelling process. Presumably, the modeller should attempt to achieve some level of realism when incorporating time and time-variant parameters into a model. That is, the treatment of time and time-variant parameters in the model should be such that they behave in a fashion analogous to the way they behave in nature. The method of choice should allow the generation of realistic statistical descriptors (mean, standard deviation, maximum, minimum, etc.). The time increment should allow at least preliminary evaluation of the transient behaviour of the system in order to examine the potential importance of transient phenomenon. The time increment used should minimize computer time while satisfying the other criteria.

Table 3.2

Summary Statistics for Radionuclide Concentration in Soil Layer 2 at 100 years

For Two Approaches to Stepping Through Time

	Mean (Bq/m^3)	Standard Deviation	Geometric Mean (Bq/m^3)	Geometric Standard Deviation	Minimum (Bq/m^3)	Maximum (Bq/m^3)
Air Concentration and Precipitation Both Uncertain Time-Variant Parameters; 1 year Increment (Layer 2, Case 1)	1.4×10^6	7.1×10^5	1.2×10^6	1.8	1.6×10^5	3.3×10^6
Air Concentration and Precipitation Both Uncertain Non-Time-Variant Paramters; 1 year Increment (Layer 2, Case 4)	1.3×10^6	1.1×10^6	9.4×10^5	2.5	4.8×10^4	6.5×10^6

There is no simple or universal answer as to how best to proceed. Consider the soil model and the information generated in the previous discussion. If we run 500 trials of 100 years duration with a time increment of 1 year, 50,000 iterations of the soil model calculation would be made. If we went to a 10 year time increment only 5000 iterations would be done. The soil model was run using the two above approaches and ran 6 times faster for the ten year time step. Should the transient behaviour shown in Figure 3.1 not be of particular interest then the ten year time step would be preferable.

In terms of generation of reliable statistical descriptors, treatment of an uncertain time-variant parameter as an uncertain non-time-variant parameter (case 4 on Table 3.1 and Table 3.2), may result in unrealistic minimums and maximums and an unreasonably dispersed sample population and is hence not recommended.

4.0 SOME TECHNIQUES FOR SENSITIVITY AND UNCERTAINTY ANALYSIS

Several techniques for sensitivity and uncertainty analysis are reviewed in general terms in this section to set the stage for the sensitivity and uncertainty analysis of the soil model, discussed in Section 5. Sensitivity analysis is employed to try to identify the most important parameters in a model independent of parameter variability. (In the classical sense sensitivity analysis indicates the rate of change of model output relative to a unit change in the various specific model inputs.) Uncertainty analysis examines output variability with respect to input variability and attempts to identify the parameters that contribute most to the uncertainty (i.e. spread) of the output distribution.

4.1 Sensitivity Index

The sensitivity index SI (Hoffman and Gardner 1983) provides an indication of sensitivity and is calculated for the ith parameter as follows:

$$SI_i = 1 - \frac{\text{Low Output Result}}{\text{High Output Result}} \qquad (4\text{-}1)$$

The model is run twice with the ith parameter at its minimum and maximum, while all other parameters are held constant at their nominal values. The closer the SI is to one (i.e. the larger the difference between the high and low dose result) the more sensitive the model is to the variable. The sensitivity index must be used with caution however, as the sensitivity results will vary depending on the nominal values chosen for the other parameters while examining the effects of one particular parameter. The limits chosen for a parameter will also affect the sensitivity index.

4.2 Graphical Presentation

One parameter is allowed to vary while all other parameters are held constant at their nominal values. The model output is then plotted against the variable of interest. If the limits of the parameter of interest are indicated on the graph, the sensitivity index can be also calculated. Non-linearities and rates of change of the output versus a particular input variable are also indicated by this technique. It must be noted that this technique only accounts for the slope at a particular point on the response surface.

4.3 Uncertainty Index

The variance of the result of a perfectly additive or multiplicative model can be estimated by combining the variances of the input parameters. For a purely additive model of the form

$$Y_a = X_1 + X_2 + X_3 .. + X_n \qquad (4\text{-}2)$$

the variance of Y_a is simply the sum of the variances of the individual parameters:

$$\mathrm{Var}(Y_a) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) \ ... + \mathrm{Var}(X_n) \qquad (4\text{-}3)$$

For a purely multiplicative model of the form:

$$Y_m = X_1 * X_2 * X_3 .. * X_n \qquad (-4\text{-}4)$$

a log transform must first be performed, which results in an additive model:

$$\ln Y_m = \ln X_1 + \ln X_2 + \ln X_3 + ... + \ln X_n \qquad (4\text{-}5)$$

The variance of $\ln Y_m$ is equal to:

$$\mathrm{Var}(\ln Y_m) = \mathrm{Var}(\ln X_1) + \mathrm{Var}(\ln X_2) + \mathrm{Var}(\ln X_3) + ... + \mathrm{Var}(\ln X_n) \qquad (4\text{-}6)$$

which is readily determined from the geometric standard deviation (GSD):

$$\mathrm{Var}\ (\ln Y_m) = [\ln\ (\mathrm{GSD}(Y_m))]^2 \qquad (4\text{-}7)$$

Since variance is a common measure of parameter uncertainty, the ratio of the variance of a particular parameter to the overall variance of the result provides an estimate of the contribution of that parameter to model uncertainty. A ratio using this concept has been termed the 'importance index' (Hoffman and Gardner 1983). However, preference has been given in this study to the use of the term "uncertainty index" to maintain consistency with the earlier definition of uncertainty analysis (see also SENES 1984).

For an additive model (as described earlier), the uncertainty index (UI_i) is given by:

$$UI_i = \frac{\mathrm{var}\ (X_i)}{\mathrm{var}\ (Y_a)} \qquad (4\text{-}8)$$

While for a multiplicative model, the uncertainty index is calculated by:

$$UI_i = \frac{\mathrm{var}\ (\ln X_i)}{\mathrm{var}\ (\ln Y_m)} = \frac{[\ln(GSD(X_i))]^2}{[\ln(GSD(Y_m))]^2} \qquad (4\text{-}9)$$

Unfortunately, this index is only strictly valid for purely additive or multiplicative models. For such models, the parameter uncertainty indices would all be less than 1.0 and would also add to 1.0. A parameter uncertainty index near 1.0 would indicate that the particular parameter contributed significantly to uncertainty in the result. Conversely, an index near zero indicates little contribution to total variation in the results and may suggest that it could be assigned a suitable 'constant' value in subsequent analyses with little loss of information.

4.4 Regression Analysis Techniques

A linear regression model is of the form:

$$Y = a+b_1X_1+b_2X_2+\ldots\ldots b_nX_n \tag{4-10}$$

where Y is the dependent variable calculated in the probabilistic assessment, the X_i's are the independent variables and b_i's are the unstandardized regression coefficients (URC's) determined by the regression analysis.

The variables X_i in the regression model are in fact the parameters to which distributions were assigned in the original soil model. In this and following sections, the term variable is used in the sense of being a variable in a regression analysis.

In essence when using regression analysis for sensitivity and uncertainty analysis, the output variable is treated as the dependent variable and the input variables as independent variables in a multiple regression analysis. If the regression model provides a good approximation of the behaviour of the assessment model, conclusions can be made relating to sensitivity and uncertainty from quantities generated in the regression.

Unfortunately the data generated in the regression analysis provides information on the model output for a particular set of input parameters. Extreme care must be exercised in generalizing the uncertainty/sensitivity information gained from regression done with one set of inputs. Different conclusions may be reached when an analysis is carried out with different inputs. A fuller discussion of the effects of different inputs on the information provided by regression analysis is contained in a recent report prepared by SENES (1985).

Various statistical indicators can be used to describe the adequacy of the linear regression model as a predictor of the assessment model's behaviour and the contribution of different input variables to output uncertainty. A fuller discussion of regression analysis including a discussion of stepwise regression is contained in Appendix A. The indicators of interest in sensitivity and uncertainty analysis are outlined below.

The R^2 value, the multiple correlation coefficient squared, provides an indicator of the regression model's ability to predict the behaviour of the dependent variable. R^2 is merely 1 minus the ratio of the square of residual error not accounted for by the regression model to the total sum of squares or:

$$R^2=1- \Sigma(Y_p-Y_i)^2/ \Sigma(\bar{Y}_i-Y)^2 \qquad (4\text{-}11)$$

where Y_p = predicted Y
Y_i = actual Y value
$\bar{Y}$ = mean value of Y

Clearly the better the model predicts Y_i the closer R^2 is to 1.

The value of the overall F statistic indicates the confidence level that can be placed on the prediction of the regression coefficients. The overall F statistic is calculated as the mean sum of squares attributable to regression divided by the mean sum of squares of the residual error (see Appendix A).

The URC's from multiple regression provide an indication of the rate of change in the output variable Y with respect to a unit change in an input variable namely $\partial Y/\partial X_i$ (eg. see Gardner 1984). It must be noted that these slopes are obtained from an assumption of a linear model. The URC's are limited in the sense that they provide an approximation of the partial derivative of output with respect to each input evaluated at the mean values of the other input variables. Should the mean values of the input variables change, the partial derivatives will change.

If we let $\bar{Y}$ and S_y denote the estimated mean and standard deviation for Y; $\bar{X}_i$ and ${}^{S}X_i$ denote the estimated mean and standard deviation for X_i (i=1,2,....,m) then,

$$\Sigma(Y_i-\bar{Y})/S_y = \Sigma\, b_{si}(X_i-\bar{X}_i)/{}^{S}X_i \qquad (4\text{-}12)$$

The b_{si}'s are called the standardized regression coefficients (SRC's) for the regression and are calculated as:

$$b_{si}=b_i{}^{S}X_i/S_y \qquad (4\text{-}13)$$

According to Helton and Iman (1982) an ordering of the absolute values of the SRC's provides a means of ranking the individual independent variables. That is the input variable whose standardized regression coefficient is largest in absolute value is deemed most important, the variable whose standardized regression coefficient is next largest in absolute value is deemed next most important, and so on. The justification for such a ranking procedure lies in the fact that the SRC's are generated so that all variables have a mean of 0 and a variance of 1. Thus, the further the standardized regression coefficient for a particular independent variable is away from 0, the more influence changes in that variable will have on the dependent variable. However, as factors such as model nonlinearities, choice of parameter distribution (i.e. a regression variable), and correlation among regression variables (input parameters) may also influence the relative importance of individual variables, a ranking of regression variables based on the standardized regression coefficients may be meaningful only when R^2 is 'large'. Since the SRC's embody both the slopes (b_i's) and the relationship between input and output variability (${}^{S}X_i/{}^{S}Y$) they are beginning to reflect a combined measure of sensitivity and uncertainty.

If the multiple regression coefficient and the overall F value indicate a good fit, the computed F value for each regression coefficient provides a good indicator of how significant a particular parameter is to the regression. The computed F value for the ith regression coefficient(F_i) is calculated as follows:

$$F_i = b_i{}^2 / SE_i{}^2 \qquad (4\text{-}14)$$

where b_i = unstandardized regression coefficient for ith parameter

SE_i= standard error for ith parameter

As suggested by equation (4-14), F_i takes account of both slope (a measure of sensitivity) and uncertainty and is thus a combined measure of sensitivity and uncertainty. The F values for each regression coefficient can be compared with tabular listings of F values for the appropriate degrees of freedom (see Appendix A) to determine the probability of a particular regression coefficient not being zero (i.e. insignificant). The F value for a regression coefficient can also be decomposed into a semipartial correlation coefficient r_{sp} (Edwards 1979) using the following equation:

$$r^2_{sp_i} = F_i \ (1-R^2)/(n-k-1) \qquad (4\text{-}15)$$

where

n = sample size

k = number of independent variables

The following statement can be made about the ith independent variable (X_i) with respect to the ith partial correlation coefficient: "Then we know that when X_i is entered last in the regression equation it accounts for r^2_{sp} of the total sum of squares over and above that accounted for by the other independent variables." (Edwards 1979)

5.0 SENSITIVITY AND UNCERTAINTY ANALYSIS OF THE SOIL MODEL

The techniques and indicators discussed in Section 4.0 were applied to the soil model simulations noted in Section 2.2. This section will present these indices and will discuss the relative merits of them. The sensitivity index, unstandardized and standardized regression coefficients, uncertainty index, F statistic and semipartial correlation coefficients from the analysis are listed in Table 5.1, for soil layer 1 at the 100th year (steady state prevails). The graphical sensitivity analysis method for three key parameters is depicted in Figures 5-1, 5-2 and 5-3. It should be noted that two multiple regressions were performed to derive the sensitivity/ uncertainty indices in Table 5.1, one on the untransformed parameters (designated UT) and one on the log-transformed parameters (designated LT). These regressions will be discussed in detail later in this section.

Since, the closer the absolute value of the indices in Table 5.1 (except for the URC's) are to the number one (1) the more "significant" the variable is, it is of interest to compare the magnitudes of the indices. Table 5.2 ranks the "significance" of the parameters according to the indices shown in Table 5.1. From examination of these rankings certain trends are apparent. These 'trends' and the relative merits of the different techniques as applied to the soil leaching model are discussed in the following paragraphs.

Table 5.1

Summary of Various Sensitivity and Uncertainty Indices for Soil Model Inputs[1]

Parameter	SI[2]	Standardized Regression Coefficient		Unstandardized Regression Coefficient		Uncertainty Index[5]	F Statistic		Semi-Partial Correlation Coefficient	
		UT[3]	LT[4]	UT	LT		UT	LT	UT	LT
Precipitation	0.88	-	-	-	-	0.07				
Air Concentration	0.99	0.59	0.56	8.6×10^7	1.00	0.30	128	1.4×10^3	0.34	0.30
Deposition Velocity	0.95	0.41	0.35	6.6×10^{-1}	0.99	0.13	59	5.5×10^2	0.16	0.12
Distribution Coefficient	0.98	0.36	0.30	1.6×10^6	0.93	0.11	46	4.1×10^2	0.12	0.09
Water Content	0.00	-0.04	0.00	-1.1×10^5	0.01	0.02	0.83	0.00	0.00	0.00
Bulk Density	0.95	0.41	0.58	1.4×10^3	1.00	0.32	63	1.5×10^3	0.17	0.33

1. Indices calculated for 100th year of simulation. Since the simulation was carried out using one year time steps, and precipitation was selected each time step, there are 100 precipitation values per trial. For this reason precipitation was not entered into the regression analyses.
2. SI - Sensitivity index
3. UT - Untransformed data used in regression: $R^2 = 0.86$.
4. LT - Log transformed data used in regression: $R^2 = 0.98$
5. Multiplicative model assumed.

Table 5.2

Ranks of Various Sensitivity and Uncertainty Indices for Soil Model Inputs

Parameter	SI[1]	Standardized Regression Coefficient UT[2]	Standardized Regression Coefficient LT[3]	Unstandardized Regression Coefficient UT	Unstandardized Regression Coefficient LT	Uncertainty Index[4]	F Statistic UT	F Statistic LT	Semi-Partial Correlation Coefficient UT	Semi-Partial Correlation Coefficient LT
Precipitation	3	-	-	-	-	5	-	-	-	-
Air Concentration	1	1	2	1	1	2	1	2	1	2
Deposition Velocity	2	2	3	5	1	3	3	3	2	3
Distribution Coefficient	1	3	4	2	2	4	4	4	3	4
Water Content	4	4	5	3	3	6	5	5	4	5
Bulk Density	2	2	1	4	1	1	2	1	2	1

1. SI - Sensivitiy index
2. UT - Untransformed data used in regression $R^2 = 0.86$.
3. LT - Log transformed data used in regression $R^2 = 0.98$
4. Multiplicative model assumed.
5. Indices were estimated for 100 years of simulation.

FIGURE 5·1

CONCENTRATION AND TIME TO STEADY STATE vs. PRECIPITATION (SOIL LAYER I)

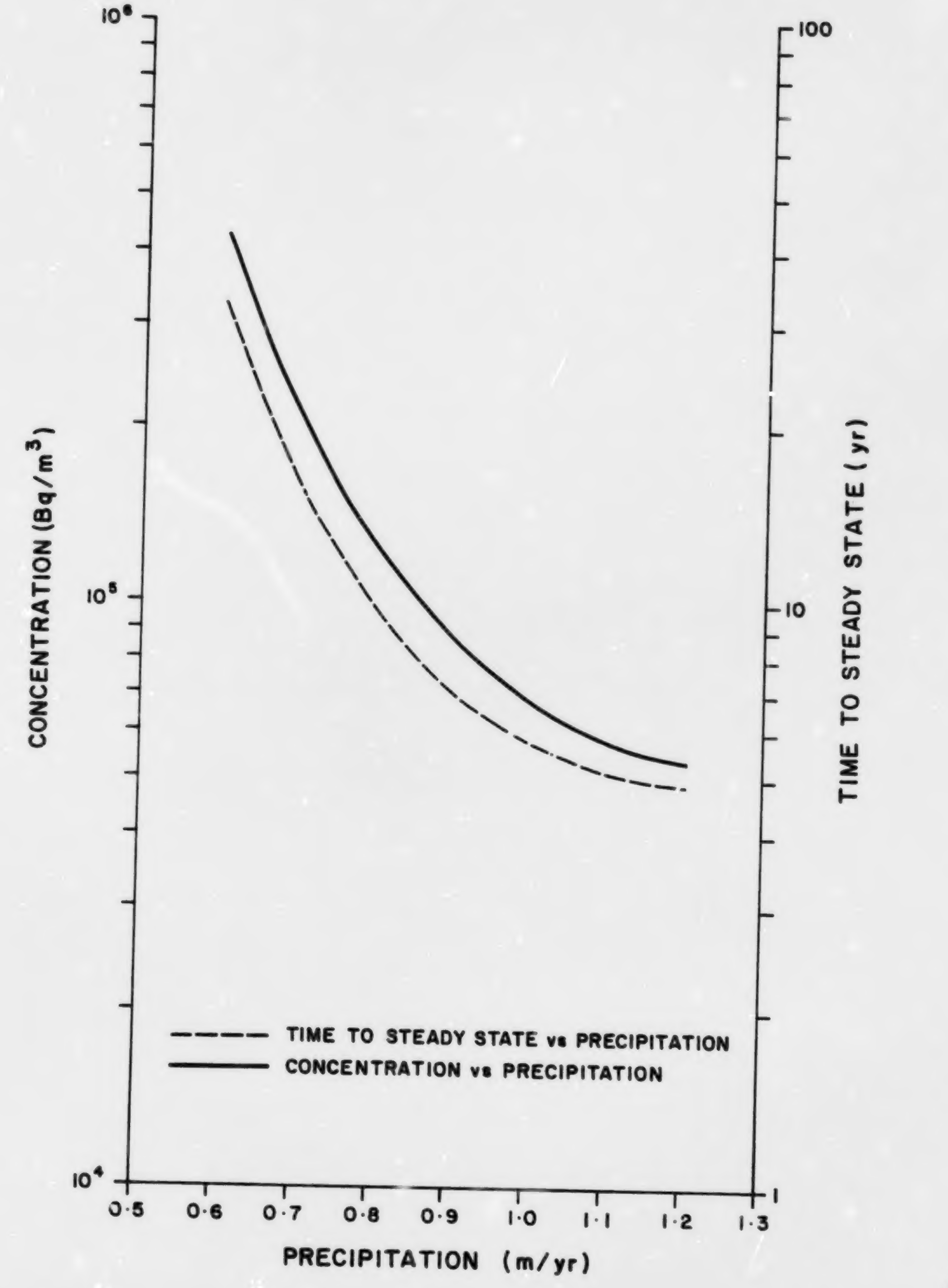


FIGURE 5·2

CONCENTRATION AND TIME
TO STEADY STATE vs. kd
(SOIL LAYER I)

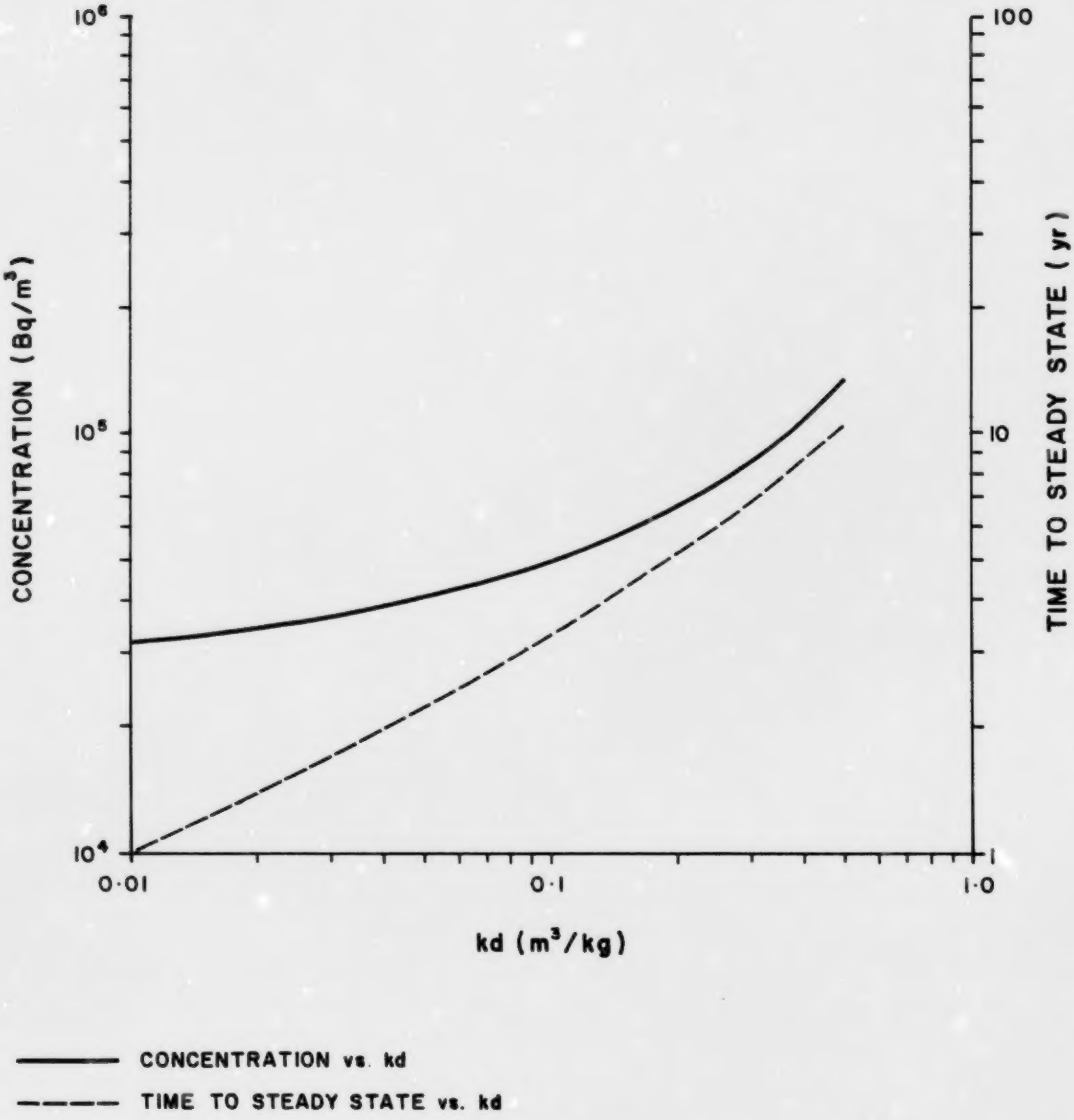
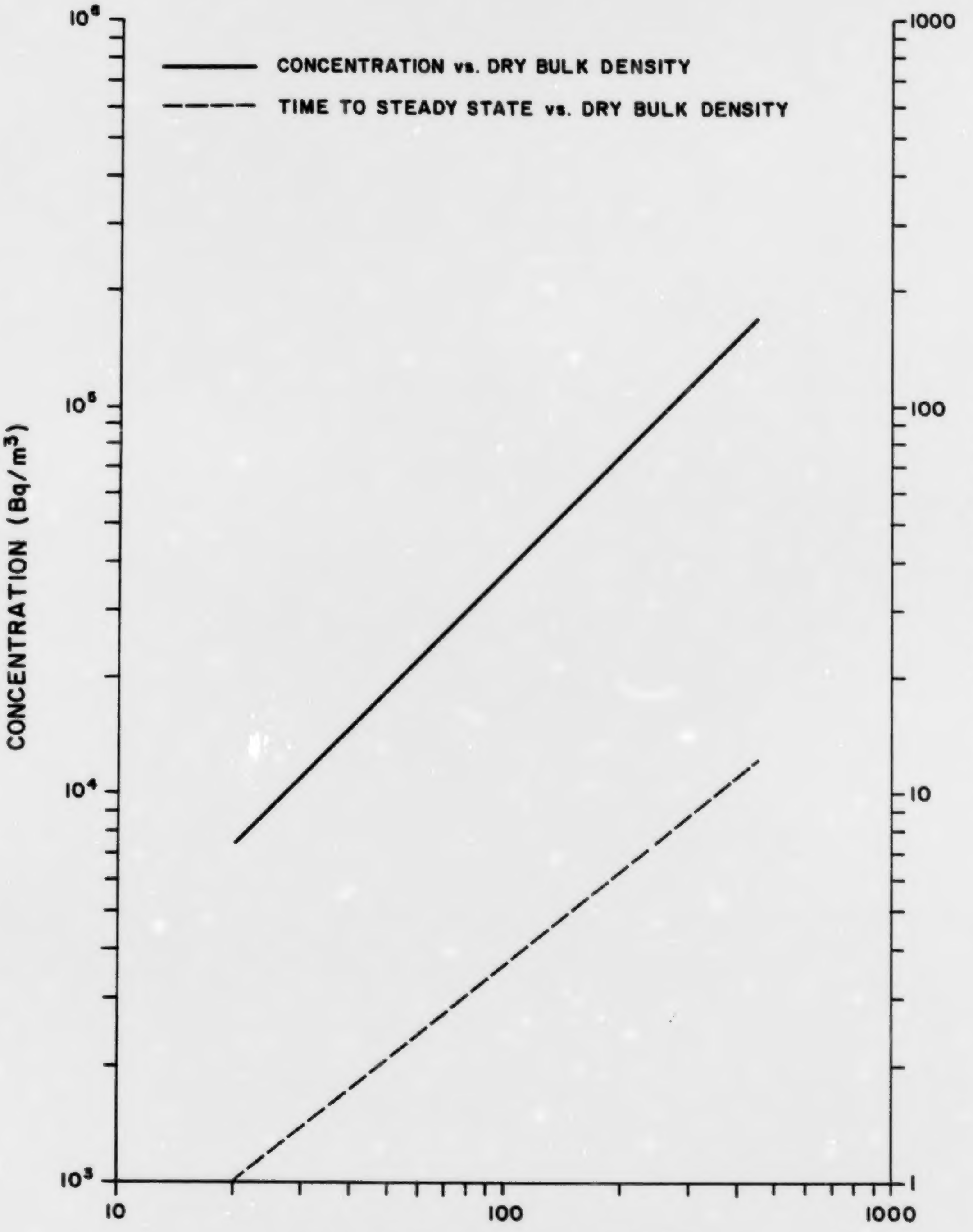


FIGURE 5·3

CONCENTRATION AND TIME TO STEADY STATE vs. DRY BULK DENSITY (SOIL LAYER I)



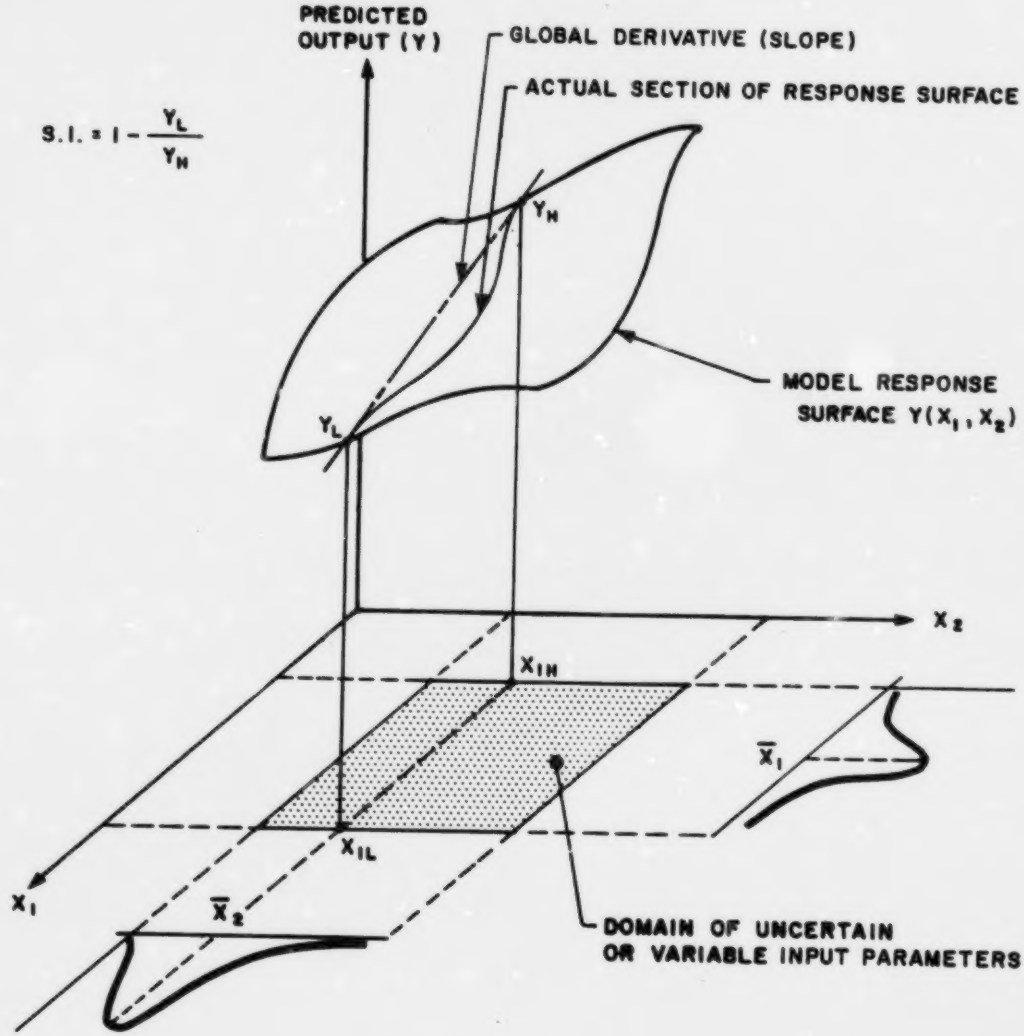
5.1 Sensitivity Index (SI) and Graphical Technique

The SI essentially assumes linear behaviour of the model over the range of the parameters being examined while all other parameters are held constant at their nominal values. This assumption is not entirely valid with the soil model as indicated by examination of the graphical technique shown in Figures 5-1, 5-2 and 5-3. These figures show that the soil system responds in a nonlinear fashion to several of the parameters even when other parameters are held constant. A simple example will help illustrate the strengths and weaknesses of the sensitivity index. Consider the two parameter model illustrated in Figure 5.4. The sensitivity index is directly related to the slope between the points Y_L and Y_H (it is the slope (global derivative) x a constant). Should the (constant) slope provide a reasonable approximation of the "actual" behaviour of this section of the response surface, the SI will prove to be a reasonable sensitivity indicator. If the "actual section of the response surface" were highly peaked or otherwise strongly contoured, the SI would not be a reliable index.

The above response surface criticism can also be directed toward the graphical technique. The graphical technique does however offer information on model nonlinearities and allows the time response of the model with respect to a particular parameter to be evaluated. The SI can also be evaluated from the graph. These two techniques can serve as initial tools to evaluate the model at the developmental (i.e. deterministic) stage of examination to make preliminary conclusions as to the significance of the various parameters. Use of parameter limits also serves as a preliminary test of the model performance under "extreme" mathematical conditions. However, if a model has a large number of parameters these techniques can become unwieldy with 2 deterministic runs and one graph required per parameter.

FIGURE 5·4

GRAPHICAL DESCRIPTION
OF SENSITIVITY INDEX
FOR PARAMETER X_1


PREDICTED OUTPUT (Y)
GLOBAL DERIVATIVE (SLOPE)
ACTUAL SECTION OF RESPONSE SURFACE
$S.I. = 1 - \frac{Y_L}{Y_H}$
Y_H
Y_L
MODEL RESPONSE SURFACE $Y(X_1, X_2)$
X_2
X_{1H}
$\overline{X}_1$
X_{1L}
X_1
$\overline{X}_2$
DOMAIN OF UNCERTAIN OR VARIABLE INPUT PARAMETERS

5.2 Uncertainty Index (UI_i)

The UI_i provides a direct comparison of input and output variance given that either the additive or multiplicative model assumption is met. The soils model is neither truly additive nor multiplicative, but rather contains a combination of additive, multiplicative and exponential factors.

Nonetheless, the uncertainty indices given on Table 5.1, which were calculated assuming a multiplicative model, add to 0.95 which closely approximates the theoretical value of 1.0 for a purely multiplicative model. The value of 0.32 for the bulk density suggests that this parameter contributes most to the variability in the predicted concentration in soil layer 1. The analysis suggests that soil bulk density is the second most sensitive contributor to the predicted concentration.

5.3 Regression Methods

It was noted earlier that untransformed (UT) and a logtransformed (LT) regression models were used to fit the soil model output and inputs. The mathematical formulation of the untransformed and logtransformed regression models are, respectively:

$$Z = a + b_1 C_A + b_2 V_g + b_3 K_d + b_4 W + b_5 \rho_s \qquad (5\text{-}1)$$

and

$$Z = a \times (C_A{**}b_1) \times (V_g{**}b_2) \times (K_d{**}b_3) \times (W{**}b_4) \times (\rho_s{**}b_5) \qquad (5\text{-}2)$$

The overall R^2 for the untransformed regression was 0.86 and for the logtransformed regression equalled 0.98. Since the regression improves in predictive ability as R^2 (the fraction of the total sum of squares due to regression) approaches unity, the regression of log-transformed values outperforms the regression on untransformed values.

Let us insert the UT and LT coefficients from the regression analysis into the respective regression equations:

$$Z = -1.1x10^{6} + (8.6x10^{7}\ C_A) + (6.6x10^{-1}\ V_g) + (1.6x10^{6}\ K_d) \\ (1.1x10^{5}W) + (1.4x10^{3}\ \rho_s) \qquad (5\text{-}3)$$

and

$$Z = 0.31xC_A{}^{1.0} \ x\ V_g{}^{0.99} \ x\ K_d{}^{0.93} \ x\ W^{0.01} \ x\ \rho_s{}^{1.0} \qquad (5\text{-}4)$$

An examination of the absolute magnitude of the URC's (b's) in either equation provides an indication of the sensitivity of the variable in the model. The URC's for the untransformed regression indicate the rate of change of the output variable (i.e. the dependent regression variable) Z with respect to a particular input parameter (independent regression variable). The URC's from the logtransformed regression do not give the rate of change of Z with respect to a particular regression variable as they do in the untransformed case (i.e. equation (5-3)), however they do indicate the mathematical proportionality between Z and a particular regression variable. Indeed in this instance there exists a readily calculable relationship between $\partial Y/\partial X_i$ and $\partial \ln Y/\partial \ln X_i$. Unfortunately with complex models calculation of the partial derivatives becomes difficult. Given that the logtransformed regression is a better fit than the untransformed regression (other transformations which serve to linearize the transformed data would lead to similar conclusions) it is not unreasonable to conclude that suitable analysis of the regression model on log transformed data should give a superior indication of sensitivity. As evidenced from Table 5.2, the rankings of the URC's for the LT model more closely parallel the rankings of the other indicators than do the URC's for the UT model.

In terms of the uncertainty indicators the standardized regression coefficients, the F statistics and the semipartial correlation coefficients for both regressions give similar rankings for the significance of the input parameters (i.e. regression variables) to the soil model. In terms of statistical rigour the F values for the URC's and the semipartial correlation coefficients derived from the statistics provide the most information in terms of the statistical significance of a variable and its contribution to overall variability. Consequently, the F statistic for the URC's (or equivalently the semi partial correlation coefficient derived from it) may be the single most useful measure of sensitivity and uncertainty. Further, if the model to be examined has a large number of variables and is to be evaluated probabilistically, the regression analysis method, if the fit is good, is not only the best sensitivity/uncertainty analysis method, it is the quickest and easiest. Its major weakness, as we ascertain it at this time, is its inability to readily accommodate uncertain time-variant parameters.

The addition of an uncertain time-variant parameter like precipitation complicates the use of regression analysis as an uncertainty/sensitivity analysis tool. The regression analyses described above were performed on soil concentration as the dependent variable and the uncertain non-time-variant parameters as independent variables. Precipitation could not be included since a single, unique value of precipitation did not exist for each trial. Rather precipitation was sampled each time step in each trial which yielded a multitude of data points (equal to the number of time steps times the number of trials). Hence, regression analysis was not found to be a suitable tool for examining time-variant parameters such as precipitation.

5.4 Discussion of Results

The sensitivity/uncertainty analysis indicates that soil water content is an unimportant parameter. Therefore little additional effort should be spent on improving the estimate for this parameter. In fact it could be treated as a constant with little loss of information.

In contrast, soil bulk density and air concentration are both important parameters. Additional effort could usefully be spent "firming up" the estimates for these parameters. Both parameters vary over several orders of magnitude. A reduction in the variance of these parameters would contribute to a reduction in the uncertainty in the dependent variable. Intuitively, air concentration and soil bulk density would seem to be important parameters since they respectively indicate the amount of incoming radionuclide and the amount of soil per unit volume capable of retaining the radionuclide.

Precipitation, according to the uncertainty and sensitivity indices is not as important as the parameters noted above. This bears out the small difference observed in the soil concentrations between the four strategies for incorporating the time variability of precipitation, (see Section 3.2). A much more significant effect was observed when the more "important" radionuclide concentration in air term was treated as an uncertain time-variant parameter (see Table 3.2)

6.0 CONCLUSIONS

As part of the Phase II probabilistic assessment study, a conceptual model was formulated to allow the buildup on soil due to deposition of airborne particulate matter to be investigated. The soil model response was then evaluated using several sensitivity and uncertainty analysis techniques to identify the relative importance of the independent input parameters and to demonstrate the utility of the techniques. The response of the soil model was also investigated using four different approaches for incorporating time-variant parameters (i.e. parameters which vary from year to year) in addition to the numerous uncertain parameters (i.e. parameters which are constant but whose true value is uncertain).

The soil model discussed in this report included two soil layers (it can, however, treat up to 5 layers); a thin exchange layer of organic matter of about 0.01 m thickness underlain by a root zone layer of about 0.15 m thickness. The change in radionuclide content with time in each of the layers was characterized as a function of deposition of particulate radioactivity from the atmosphere on layer 1 (or leachate inflow from layer 1 to layer 2). The soil model included eight input parameters requiring assignment of probability distributions. The parameters included in the model were total precipitation, evapotranspiration losses, particulate deposition velocity, the radioactive decay constant for radium, the radium content of atmospheric particulate matter, the radium distribution coefficient between solid phase and liquid phases,the soil bulk density and the soil moisture content.

The response of the soil model was evaluated using four different approaches to handling precipitation which is a truly time-variant parameter. The approaches were: i) hold precipitation constant in all trials at the assessed mean annual value; ii) treat precipitation as one of the uncertain parameters and sample from a distribution of possible values at the beginning of each trial; iii) run the probabilistic analysis using ten year time steps and sample precipitation at each time increment from a distribution characteristic of ten year mean values; and iv) run the probabilistic analysis using annual time steps and sample precipitation each time step from a distribution of annual mean values.

The results of this analysis on layer 1 at year 100 when steady-state conditions prevail showed that all four approaches produced nearly identical results for major statistical descriptors of soil radium concentration. The mean values and standard deviations calculated for the soil radium concentration by all four approaches were essentially equal. However, in an additional run treating both precipitation and radium concentration in air as time-variant parameters only the mean value was similar to the non-time-variant parameter approach. The non-time-variant approach yielded a more dispersed population with more extreme minimum and maximum values.

The soil concentration in layer 1 showed marked transient behaviour when precipitation was allowed to vary in time whereas the effects of the time-variant parameters were totally obscurred in layer 2. Based on the soil model analysis it is concluded that the ability to model time-variant parameters provides a more realistic representation of real system behaviour. The size of the time step selected for any application must balance computing time requirements against the detail desired in the output results.

Analysis of the results from running the soil model in a probabilistic mode was attempted using several different statistical techniques. The sensitivity and uncertainty methods applied in the study ranged from the fairly simplistic to the more sophisticated and included: the sensitivity index, a graphical presentation, the uncertainty index and regression analysis.

Interestingly, all methods tested yielded fairly comparable interpretations of the relative contributions of the input parameters to model sensitivity and uncertainty even though the conditions for applying the simpler methods (i.e. the sensitivity and uncertainty indices) were not fully met. Regression analysis provided the most rigorous basis for interpreting the predicted soil concentration data although it failed to account for time-varying parameters. Regression analysis was applied to both the untransformed data set and to a log-transformed data set. In both cases, all statistical measures of the goodness of fit of the regression equations were significant although regression of the log-transformed data was better. Of the six input parameters treated as uncertain and/or time-variant inputs, only the water content of the soil was found to contribute little to the sensitivity and uncertainty of the model predictions.

REFERENCES

Ang, A.H.S., and Wilson, H.T., 1975. "Probability Concepts in Engineering Planning and Design." John Wiley and Sons Inc., New York.

Draper, V., and Smith, H., 1981. "Applied Regression Analysis, Second Edition." John Wiley & Sons Inc., New York.

Dixon, W.J., 1977 (Ed.). "BMD-Biomedical Computer Programs." University of California Press, Berkley, California.

Edwards, A.L., 1979. "Multiple Regression and the Analysis of Variance and Covariance." W.H. Freeman and Company, New York.

Gardner, R.H., Trabalka, J.R., and Emanuel, W.R., 1984. "Methods of Uncertainty Analysis for a Global Carbon Dioxide Model." Environmental Sciences Division, ORNL, (Draft Paper).

Helton, J.C., and Iman, R.L., 1982. "Sensitivity Analysis of a Model for Environmental Movement of Radionuclides." Health Physics, Vol. 42, No. 5.

Hoffman, F.O. and Gardner, R.H., 1983. "Evaluation of Uncertainties in Radiological Assessment Models." In Radiological Assessment: A Textbook on Environmental Dose Analysis, J.E. Till and H.R. Meyers (eds), U.S. Nuclear Regulatory Commission, Washington, D.C., NUREG/CR-3332, ORNL-5968.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long Term Effects of Uranium Mill Tailings in Canada-Phase I." Research Report Prepared for the National Uranium Tailings Program, CANMET, EMR, Ottawa, Under Supply and Services Canada Contract Number OSQ83-00236.

SENES Consultants Limited, 1985. "Uncertainty Analysis and Probabilistic Modeling." Research Report Prepared for the Atomic Energy Control Board, Ottawa, Under Supply and Services Canada Contract Number OSQ84-00285.

APPENDIX A: SOME NOTES ON REGRESSION ANALYSIS TECHNIQUES

A.1 Multiple Linear Regression

This appendix details some of the considerations of multiple linear regression analysis using the Biomedical Computer Programs (BMD) (Dixon, 1977). The multiple regression equation fitted by REGRESS using the method of least squares with n different X variables is:

$$Y = a + b_1X_1+b_2X_2 + \dots b_nX_n$$

The total sum of squares for the Y variable can be partitioned into two components or:

$$SS_{tot} = SS_{reg} + SS_{res}$$

$$\Sigma(Y_i-\bar{Y})^2 = \Sigma(Y_p-\bar{Y})^2 + \Sigma(Y_p-Y_i)^2$$

where

Y_i = observed value of Y

$\bar{Y}$ = mean value of Y

Y_p = regression predicted value of Y

The object of the regression is to minimize the residual sum of squares $\Sigma(Y_p-Y_i)^2$. The squared multiple correlation coefficient (R^2) embodies this concept.

$$R^2 = \frac{SS_{reg}}{SS_{tot}}$$

The larger R^2, the smaller SS_{res} and the better the fit.

A description follows of the output printout of REGRESS shown in Table A1 for a regression on the untransformed soil model variables. Referring to the block numbers given on the left hand margin in Table A1:

1. First and last line of input data - used as a check

2. Basic statistics on variables - mean ($\bar{X}$), standard deviation (S), minimum and maximum

3. Variance-covariance matrix - the diagonal contains the variance of each variable. The numbers below the diagonal contains the covariance of each variable. The covariance of variable X and Y (C_{xy}) is:

$$C_{xy} = \frac{\Sigma(X_i-\bar{X})\ (Y_i-\bar{Y})}{n-1}$$

   From Ang and Wilson (1975) if C_{xy} is large and positive the values of X and Y tend to be both large or both small relative to their respective means. If C_{xy} is large and negative the values of X tend to be large when the values of Y are small and vice versa, relative to their respective means. If C_{xy} is small or zero there is little or no (linear) relationship between the value of X and Y. Therefore C_{xy} is a measure of the linear interrelationship between the variables X and Y.

4. For the purposes of establishing linear interrelationships between variables it is preferable to use the normalized covariance or correlation coefficient (r) which is defined as:

$$r = \frac{C_{xy}}{S_X\ S_Y}$$

The correlation matrix lists the correlation coefficient between the paired variables. These correlation coefficients do not embody multiple correlations.

When $r = \underline{+}1$, variables X and Y are linearily related and all of the points fall on the regression line. When $r = 0$ values of X and Y are not linearily related (they may however be functionally related).

5. For each dependent variable the mean, standard deviation and correlation coefficient (X_i with Y) is shown. The unstandardized regression coefficients (URC's) are presented next. The URC's indicate, the rate of change of Y with respect to each of the independent variables (X_i's). The standardized regression coefficients are then shown. These are calculated before the unstandardized coefficients and are actually a computing "trick" to minimize the size of numbers and subsequent roundoff problems. The standardized regression coefficients are calculated by standardizing the X_i variables with a mean of zero and a standard deviation of 1. The unstandardized regression coefficient (b_i) can be obtained from the standardized coefficient (b_{si}) as follows:

$$b_i = b_{si} \frac{S_Y}{S_{X_i}}$$

The standard error (SE_i) of b_i is calculated as:

$$SE_i = \left[\frac{MS_{res}}{\Sigma(X_i - \bar{X})^2 (1 - R_i{}^2)} \right]^{1/2}$$

where $R_i{}^2$ is the squared multiple correlation with the remaining X_i variables and MS_{res} is the mean square deviation from regression shown in block 7 of Table A1.

Stated verbally the standard error is the square root of the total residual mean squares divided by the total variance attributable purely to X_i.

For a sample of size n with k independent variables the standard error of the regression coefficients provides an estimate of the confidence interval for the respective regression coefficients as follows:

$$b_i \pm t\ (n-k-1,\ 1-\alpha\ 2)\ SE_i$$

where t is the area under the two tails of the t distribution and α is 1 - the confidence interval. For example using the SE_i of variable number 5 from a table of percentage points of the t-distribution, the 95% confidence interval for b_5 is:

$$1463 \pm t(94,1-0.05/2) \times 185 = 1463 \pm 1.99 \times 185$$
$$= 1094 \text{ to } 1831$$

The F value is computed as:

$$F = \frac{b^2{}_i}{SE_i{}^2}$$

The computed F tells us how significant a variable is in the regression. The higher the F value relative to other values the more important that variable. The computed F value can be compared to the F distribution points in a table of percentage points of the F-distribution to establish how good a predictor the regression coefficient is. For example, using variable 2 from Table A1 and comparing the computed F of 59.4 with F(5,94,0.99) = 3.24, the computed F is higher. Hence rejection of the hypothesis that b_2 equals zero can be made with a risk of less than 1% of being wrong (Draper and Smith 1981).

The F value can also be used in calculating the semipartial correlation coefficient. The semipartial correlation coefficient is the correlation between X_i and Y after the the variance that X_1 had in common with the other X_i's has been removed from X_1 (Edwards 1979). This tells us that when X_1 is entered last in the regression equation it accounts for the semipartial correlation coefficient squared of the total sum of squares over and above that already accounted for by the other X_i's. The semipartial correlation (r_{spi}) coefficient for the X_ith is calculated as:

$$(r_{spi})^2 = F_i \ (1-R^2)/(n-k-1)$$

For example, variable number 2 has a $r^2{}_{sp2}$ of:

$$\begin{aligned}(r_{sp2})^2 &= 59.4 \ (1-0.74)/94 \\ &= 0.16\end{aligned}$$

It can be seen that the calculation of the r_{sp} for a particular variable in the regression is merely the F statistic times a constant. Since we are primarily interested in the relative influence of the various independent variables (X_i's) on the dependent variable (Y) the F statistic and the r_{sp} are the best indicators of this.

6. Block 6 details the Y intercept, the standard error of the intercept, the multiple correlation coefficient (R) and R^2. The closer R^2 is to one the better the model fits the data. The standard error (SE_e) of the estimate is calculated as:

$$SE_e = (\text{mean squares deviation from regression})^{1/2}$$

7. Analysis of variance for the regression. This analysis helps indicate the degree of precision that can be attached to the estimate of the regression line. The sum of squares about the mean can be divided into 2 sections (Draper and Smith 1981).

$$\begin{bmatrix}\text{sum of squares}\\ \text{about the mean}\end{bmatrix} = \begin{bmatrix}\text{sum of squares}\\ \text{attributable to}\\ \text{regression}\end{bmatrix} + \begin{bmatrix}\text{residual sum of}\\ \text{squares not attributable}\\ \text{to regression}\end{bmatrix}$$

$$\Sigma(Y_i-\bar{Y})^2 = \Sigma(Y_p-\bar{Y})^2 + \Sigma(Y_i-Y_p)^2$$

This shows that, of the variation in the Y's about their mean, some of the variation can be ascribed to the regression line and some to the fact that the actual observations do not all lie on the regression line. If they did, the sum of squares attributable to the regression would be zero. From this procedure we can see that a way of assessing how useful the regression line will be as a predictor is to see how much of the sum of squares about the mean has fallen into the sum of squares due to regression and how much into the sum of squares about regression. We shall be pleased if the sum of squares attributable to regression is much greater than the sum of squares about regression. R^2 embodies this concept and is the sum of squares attributable to regression divided by the sum of squares about the mean (Draper and Smith 1981).

The mean square is the sum of squares divided by the appropriate degress of freedom. The F value is the mean squares attributable to regression divided by the mean squared deviation from regression. It serves as an indicator of how good a predictor the regression equation is. If the calculated F is greater than the $F(5,94,0.99)=3.24$, the regression equation is a good predictor. Essentially we are rejecting the hypothesis that the

URC's are zero and running a risk of less than 1% of being wrong. According to Draper and Smith (1981) an equation should be regarded as a good predictor if the F statistic exceeds the selected percentage point of the F distribution by 4 times. Our regression satisfies this criteria.

A.2 The Stepwise Regression Procedure

The following discussion was excerpted from Draper and Smith (1981) (See Table A.3 for computer output of stepwise regression).

The procedure involves the re-examination at every stage of the regression of the variables incorporated into the model in previous stages. A variable which may have been the best single variable to enter at an early stage may, at a later stage, be superfluous because of the relationships between it and other variables now in the regression. To check on this, the partial F criterion for each variable in the regression at any stage of calculation is evaluated and compared with a pre-selected percentage point of the appropriate F distribution. This provides a judgment on the contribution made by each variable as though it had been the most recent variable entered, irrespective of its actual point of entry into the model. Any variable which provides a nonsignificant contribution is removed from the model. This process is continued until no more variables will be admitted to the equation and no more are rejected.

STEP 1. The stepwise procedure starts with the simple correlation matrix and enters into regression the X variable most highly correlated with the response.

STEP 2. Using the partial correlation coefficients as before, it now selects, as the next variable to enter regression, that X variable whose partial correlation with the response is the highest of the remaining variables.

STEP 3. The method now examines the contribution this next X would have made if it has been entered first and the first X was entered second. The value of the partial F is checked against the pre-selected percentage point of the F distribution to see if this variable should be retained. The stepwise method now selects as the next variable to enter, the one most highly partially correlated with the response (given that the first variable is already in regression).

STEP 4. A regression equation is now determined by least squares. At this point, partial F-tests for the first two are made to determine if they should remain in the regression equation.

STEP 5. The procedure continues for the remaining variables.

TABLE A1
MULTIPLE LINEAR REGRESSION EXAMPLE
REGRESSION OF SOIL RADON CONCENTRATION VS. AIR CONCENTRATION, DEPOSITION VELOCITY, Kd, WATER CONTENT AND BULK DENSITY

```
PROGRAM REGRESS - MULTIPLE LINEAR REGRESSION


      DATA FOR RECORD:      1
1        .40316D-02  .28978D+06  .20662D+00  .11998D+01  .22364D+03  .93440D+05
      DATA FOR RECORD:    100
       .28266D-02  .79696D+06  .21129D+00  .89546D+00  .51252D+02  .62612D+05

      VARIABLE     MEAN   STAND. DEV.      MINIMUM       MAXIMUM

           1        .00513        .00299
           2 588430.64000 272616.18523         .00078         .01448 - AIR CONCENTRATION
2          3        .26229        .09798 108209.00000 1262460.00000 - DEPOSITION VELOCITY
           4       1.06316        .18442         .06682         .46092 - Kd
           5     243.69320     122.73025         .75668        1.39399 - WATER CONTENT
           6 389419.16600 435380.19697       23.46640      446.12500 - BULK DENSITY
                                          13455.80000 2117380.00000 - CONCENTRATION IN LAYER 1

      VARIANCE-COVARIANCE MATRIX

          1    .89663D-05
          2   -.87762D+02  .74320D+11
3         3    .76475D-05 -.52715D+04  .96005D-02
          4   -.33971D-04  .47255D+04  .13344D-02  .34012D-01
          5   -.24176D-02  .18681D+07  .33882D+00  .49416D+00  .15063D+05
          6    .72815D+03  .35282D+11  .12933D+05 -.79478D+03  .23547D+08  .18956D+12
                    1           2           3           4           5           6
      CORRELATION MATRIX

          1    1.00000
          2   -.10751 1.00000
4         3    .02607 -.19735 1.00000
          4   -.06152  .09399  .07384 1.00000
          5   -.00658  .05583  .02818  .02183 1.00000
          6    .55853  .29726  .30317 -.00990  .44066 1.00000
                  1       2       3       4       5       6
```

Table A.1 (continued)

```
 PROBLEM TITLE:   CONCENTRATION LAYER 1 T=100 YEARS VS CAIR,DEPVEL,KD,WC,RHO

                                                   UNSTANDARDIZED STANDARDIZED
      VARIABLE      MEAN       STANDARD      CORRELATION    REGRESSION    REGRESSION    STD. ERROR      COMPUTED
       NO.                     DEVIATION       X VS Y       COEFFICIENT   COEFFICIENT   OF REG.COEF.   F VALUE
        1        .00513          .00299       .5585386274406.74395          .59336 7610850.45558     128.49836
5       2   588430.64000   272616.18523       .29726           .66092       .41384        .08576      59.39344
        3        .26229          .09798       .30317 1605417.81312          .36130  236802.67286      45.96243
        4       1.06316          .18442      -.00990 -113260.26819         -.04798  123968.95141        .83470
        5     243.69320      122.73025        .44066     1462.72356         .41233     184.80710      62.64512
     DEPENDENT
        6   389419.16600   435380.19697

     INTERCEPT              -1098912.38914
     STD. ERROR OF INTERCEPT162552.07334
     MULTIPLE CORRELATION (R)      .86379
6    R-SQUARED  (R*R)              .74613
     STD. ERROR OF ESTIMATE 225125.63622
     DETERMINANT OF MATRIX         .92596
     NUMBER OF OBSERVATIONS           100

                        ANALYSIS OF VARIANCE FOR THE REGRESSION
        SOURCE OF VARIATION       DEGREES        SUM OF          MEAN               F VALUE
                                  OF FREEDOM     SQUARES         SQUARES
7    ATTRIBUTABLE TC REGRESSION      514001969780190.380 2800393956038.077          55.25470
     DEVIATION FROM REGRESSION     94 4764065895682.609   50681552081.730
          TOTAL                      9918766035675872.990
```

Table A2
Multiple Linear Regression on Logtransformed Parameters

```
 FORMAT OF DATA:  (7f10.7)

 DATA TREATED AS HAVING NO MISSING VALUES

0DATA FOR RECORD:      1
   .46609D+01  .60810D-02 -.18964D+01  .55364D+01 -.68486D+00  .12073D-01  .14918D+01
0DATA FOR RECORD:    100
   .59994D+01 -.52300D-02 -.22049D+01  .60833D+01 -.67761D+00  .91857D-01  .25658D+01

0 VARIABLE     MEAN    STAND. DEV.     MINIMUM       MAXIMUM

      1       5.37321      .49358      4.24008       6.64045
      2       -.04192      .06700      -.21875        .07824
      3      -2.31514      .29338     -3.27743      -1.84357
      4       5.75839      .21989      5.12988       6.12054
      5       -.62301      .20641     -1.30945       -.31454
      6        .03816      .07744      -.11701        .14586
      7       2.24866      .35411      1.34678       2.65068

0
0VARIANCE-COVARIANCE MATRIX
0
    1      .24362D+00
    2    -.68590D-02  .44890D-02
    3     .56026D-01  .11817D-02  .86069D-01
    4     .46775D-01 -.20178D-02 -.47412D-02  .48352D-01
    5     .74981D-02  .25030D-02 -.15875D-01 -.85136D-02  .42604D-01
    6     .30939D-02  .79383D-03  .55429D-03  .21242D-02  .89722D-03  .59965D-02
    7     .12034D+00 -.54186D-04 -.73572D-02  .78097D-02 -.56345D-02  .91389D-03  .12540D+00
0              1           2           3           4           5           6           7
0
0CORRELATION MATRIX
0
    1    1.00000
    2    -.20741 1.00000
    3     .38691  .06012 1.00000
    4     .43098 -.13696 -.07350 1.00000
    5     .07360  .18099 -.26216 -.18758 1.00000
    6     .08095  .15300  .02440  .12475  .05613 1.00000
    7     .68854 -.00228 -.07082  .10030 -.07709  .03333 1.00000
0           1       2       3       4       5       6       7
1PROBLEM TITLE:    LOG Z AS F(LOG(PRECIP,CAIR,VD,KD,WC,BD))

                                                        UNSTANDARDIZED STANDARDIZED
VARIABLE       MEAN      STANDARD      CORRELATION     REGRESSION     REGRESSION     STD. ERROR      COMPUTED
  NO.                    DEVIATION       X VS Y        COEFFICIENT    COEFFICIENT   OF REG.COEF.     F VALUE
   2         -.04192       .06700        -.20741        -1.88587        -.25600         .01897     9882.21663
   3        -2.31514       .29338         .38691          .99994         .59435         .00439    51839.37614
   4         5.75839       .21989         .43098          .99890         .44502         .00582    29506.58252
   5         -.62301       .20641         .07360          .99081         .41434         .00642    23825.40575
   6          .03816       .07744         .08095          .01870         .00293         .01618        1.33566
   7         2.24866       .35411         .68854          .99974         .71725         .00349    82242.34513
```

Table A2 Continued

```
 DEPENDENT
    1        5.37321        .49358

0INTERCEPT                       .22559
0STD. ERROR OF INTERCEPT         .03355
0MULTIPLE CORRELATION (R)        .99972
0R-SQUARED  (R*R)                .99943
0STD. ERROR OF ESTIMATE          .01215
0DETERMINANT OF MATRIX           .77690
0NUMBER OF OBSERVATIONS             100
0
                      ANALYSIS OF VARIANCE FOR THE REGRESSION
0    SOURCE OF VARIATION        DEGREES         SUM OF          MEAN              F VALUE
                               OF FREEDOM      SQUARES         SQUARES
 ATTRIBUTABLE TO REGRESSION        6           24.105           4.017         27233.25312
 DEVIATION FROM REGRESSION        93             .014            .000
      TOTAL                       99           24.118
```

Table A.3
Stepwise Regression on the Logtransformed Parameters

```
NUMBER OF CASES                    100
NUMBER OF VARIABLES                  7
NUMBER OF SUB-PROBLEMS               1


SUB-PROBLM    1
 DEPENDENT VARIABLE               1
 MAXIMUM NUMBER OF STEPS         14
 F-LEVEL FOR INCLUSION      .010000
 F-LEVEL FOR DELETION       .005000
 TOLERANCE LEVEL            .001000


STEP NUMBER     1
VARIABLE ENTERED     7

MULTIPLE R                 .6885
STD. ERROR OF EST.         .3598

ANALYSIS OF VARIANCE
                        DF    SUM OF SQUARES    MEAN SQUARE    F-RATIO
        REGRESSION       1           11.4340        11.4340    88.3403
        RESIDUAL        98           12.6843          .1294

                       VARIABLES IN EQUATION                             .
                STANDARDIZED  UNSTANDARDIZED                             .
    VARIABLE     COEFFICIENT     COEFFICIENT  STD. ERROR  F TO REMOVE    .                    VARIABLES NOT IN EQUATION
                                                                         .                     PARTIAL
                                                                         .      VARIABLE        CORR.       TOLERANCE     F TO ENTER
     (CONSTANT       3.21514 )                                           .
    BD      7         .68854          .95971      .10211  88.3403 (2)    .
                                                                         .      PRECIP  2     -.28383      1.0000          8.4993 (2)
STEP NUMBER     2                                                        .      CAIR    3      .60227       .9950         55.2106 (2)
VARIABLE ENTERED     3                                                   .      DEPVEL  4      .50159       .9899         32.6083 (2)

MULTIPLE R                 .8154
STD. ERROR OF EST.         .2887

ANALYSIS OF VARIANCE
                        DF    SUM OF SQUARES    MEAN SQUARE    F-RATIO
        REGRESSION       2           16.0350         8.0175    96.2091
        RESIDUAL        97            8.0834          .0833
```

Table A.3 continued

```
                       VARIABLES IN EQUATION                          .                 VARIABLES NOT IN EQUATION
              STANDARDIZED  UNSTANDARDIZED                            .                 PARTIAL
 VARIABLE      COEFFICIENT   COEFFICIENT   STD. ERROR    F TO REMOVE  .   VARIABLE      CORR.      TOLERANCE     F TO ENTER
                                                                      .
                                                                      .
  (CONSTANT        4.82345 )                                          .
  CAIR     3        .43787        .73667      .09914     55.2107 (2)  .   PRECIP  2   -.40162       .9964         18.4631 (2)
  BD       7        .71954       1.00293      .08214    149.0929 (2)  .   DEPVEL  4    .68032       .9855         82.7157 (2)
                                                                      .   KD      5    .43866       .9221         22.8736 (2)

 STEP NUMBER    3
 VARIABLE ENTERED    4

 MULTIPLE R                .9055
 STD. ERROR OF EST.        .2127

 ANALYSIS OF VARIANCE
                      DF    SUM OF SQUARES     MEAN SQUARE     F-RATIO
      REGRESSION       3          19.7762          6.5921    145.7442
      RESIDUAL        96           4.3421           .0452

                       VARIABLES IN EQUATION                          .                 VARIABLES NOT IN EQUATION
              STANDARDIZED  UNSTANDARDIZED                            .                 PARTIAL
 VARIABLE      COEFFICIENT   COEFFICIENT   STD. ERROR    F TO REMOVE  .   VARIABLE      CORR.      TOLERANCE     F TO ENTER
                                                                      .
                                                                      .
  (CONSTANT        -.08267 )                                          .
  CAIR     3        .46434        .78121      .07320    113.8818 (2)  .   PRECIP  2   -.42750       .9785         21.2444 (2)
  DEPVEL   4        .39674        .89054      .09792     82.7157 (2)  .   KD      5    .80860       .8824        179.4344 (2)
  BD       7        .68163        .95008      .06079    244.2534 (2)  .   WC      6   -.00617       .9828           .0036 (2)

 STEP NUMBER    4
 VARIABLE ENTERED    5

 MULTIPLE R                .9683
 STD. ERROR OF EST.        .1258

 ANALYSIS OF VARIANCE
                      DF    SUM OF SQUARES     MEAN SQUARE     F-RATIO
      REGRESSION       4          22.6152          5.6538    357.3369
      RESIDUAL        95           1.5031           .0158

                       VARIABLES IN EQUATION                          .                 VARIABLES NOT IN EQUATION
              STANDARDIZED  UNSTANDARDIZED                            .                 PARTIAL
 VARIABLE      COEFFICIENT   COEFFICIENT   STD. ERROR    F TO REMOVE  .   VARIABLE      CORR.      TOLERANCE     F TO ENTER
                                                                      .
                                                                      .
  (CONSTANT        -.17203 )                                          .
  CAIR     3        .56746        .95471      .04519    446.2772 (2)  .   PRECIP  2   -.99536       .9456      10059.3400 (2)
  DEPVEL   4        .47001       1.05500      .05920    317.5872 (2)  .   WC      6   -.14490       .9735          2.0159 (2)
  KD       5        .36524        .87338      .06520    179.4344 (2)  .
  BD       7        .70974        .98926      .03607    752.0454 (2)  .
```

Table A.3 continued

```
   STEP NUMBER    5
   VARIABLE ENTERED    2

   MULTIPLE R                 .9997
   STD. ERROR OF EST.         .0122

   ANALYSIS OF VARIANCE
                        DF    SUM OF SQUARES    MEAN SQUARE    F-RATIO
          REGRESSION     5          24.1044          4.8209 32564.7900
          RESIDUAL      94            .0139           .0001

                  VARIABLES IN EQUATION                                 .
             STANDARDIZED  UNSTANDARDIZED                               .           VARIABLES NOT IN EQUATION
    VARIABLE   COEFFICIENT    COEFFICIENT  STD. ERROR    F TO REMOVE   .                  PARTIAL
                                                                        .   VARIABLE      CORR.      TOLERANCE    F TO ENTER
                                                                        .
     (CONSTANT          .22096 )                                        .
    PRECIP  2          -.25553     -1.88243     .01877   10059.3300 (2) .   WC      6   .11899      .9495          1.3357 (2)
    CAIR    3           .59449      1.00018     .00439   51791.3000 (2) .
    DEPVEL  4           .44549       .99996     .00575   30215.9600 (2) .
    KD      5           .41456       .99132     .00642   23876.3900 (2) .
    BD      7           .71733       .99985     .00349   82030.1600 (2) .

   STEP NUMBER    6
   VARIABLE ENTERED    6

   MULTIPLE R                 .9997
   STD. ERROR OF EST.         .0121

   ANALYSIS OF VARIANCE
                        DF    SUM OF SQUARES    MEAN SQUARE    F-RATIO
          REGRESSION     6          24.1046          4.0174 27234.4600
          RESIDUAL      93            .0137           .0001

                  VARIABLES IN EQUATION                                 .
             STANDARDIZED  UNSTANDARDIZED                               .           VARIABLES NOT IN EQUATION
    VARIABLE   COEFFICIENT    COEFFICIENT  STD. ERROR    F TO REMOVE   .                  PARTIAL
                                                                        .   VARIABLE      CORR.      TOLERANCE    F TO ENTER
                                                                        .
     (CONSTANT          .22558 )                                        .
    PRECIP  2          -.25600     -1.88587     .01897    9882.6530 (2) .
    CAIR    3           .59435       .99994     .00439   51841.6800 (2) .
    DEPVEL  4           .44502       .99890     .00582   29507.9000 (2) .
    KD      5           .41434       .99081     .00642   23826.4700 (2) .
    WC      6           .00293       .01870     .01618       1.3357 (2) .
    BD      7           .71725       .99974     .00349   82246.0100 (2) .
OF-LEVEL OR TOLERANCE INSUFFICIENT FOR FURTHER COMPUTATION
```

Table A.3 continued

```
0  SUMMARY TABLE

   STEP            VARIABLE            MULTIPLE                INCREASE      F VALUE TO         SUM OF SQUARES
  NUMBER      ENTERED  REMOVED      R             RSQ          IN RSQ     ENTER OR REMOVE          ADDED

     1        BD       7          .6885         .4741          .4741          88.3403             11.4340
     2        CAIR     3          .8154         .6648          .1908          55.2107              4.6009
     3        DEPVEL   4          .9055         .8200          .1551          82.7157              3.7413
     4        KD       5          .9683         .9377          .1177         179.4344              2.8390
     5        PRECIP   2          .9997         .9994          .0617       10059.3300              1.4892
     6        WC       6          .9997         .9994          .0000           1.3357               .0002
 PROGRAM TERMINATED
```

DISCUSSION PAPER ON

TECHNIQUES FOR HANDLING CORRELATED PARAMETERS

by

Terry Andres and Bruce Goodwin
Atomic Energy of Canada Limited
Whiteshell Nuclear Research Establishment
Pinawa, Manitoba

for

SENES Consultants Limited
Willowdale, Ontario

June 1985

Abstract

This paper opens with a discussion of the need to consider parameter correlations in simulation models to achieve credibility and acceptability. Correlations are imposed to mark relationships which exist in a natural system. Methods for ascertaining parameter correlations through basic research and eliciting expert opinions are examined. Next, three techniques for generating correlated parameter values in a probabilistic analysis are reviewed. The preferred method involves the construction of functional relationships where the form of the relationship between parameters is known or suspected. Another approach is to use random number generation techniques that enforce correlations. Two such techniques are presented.

TABLE OF CONTENTS

Page

Abstract i

1.0 INTRODUCTION 1

2.0 A CORRELATION EXAMPLE 3

3.0 REASONS FOR CORRELATING PARAMETER VALUES 5
- 3.1 General 5
- 3.2 Use of Available Scientific Data and Expert Opinion 5
- 3.3 Avoiding Unlikely Simulation Scenarios 6
- 3.4 Improved Sensitivity Analysis 6

4.0 METHODS OF ASCERTAINING PARAMETER CORRELATIONS 7
- 4.1 General 7
- 4.2 Research Program 7
- 4.3 Eliciting Expert Opinion 8
- 4.4 Multi-Variate Correlations 9

5.0 METHODS OF GENERATING CORRELATED PARAMETER VALUES 11
- 5.1 General 11
- 5.2 Constructing Functional Relationship 11
- 5.3 Kleijnen's Correlated Generation Technique 14
- 5.4 Rank Correlations by Iman and Conover's Technique 16

6.0 CONCLUSION 17

REFERENCES 19

LIST OF FIGURES

		Follows Page
Figure 1	Scatterplots Showing Different Levels of Correlation	1
Figure 2	Reasons for Correlating Parameter Values	5
Figure 3	Possible Meanings of "A is Correlated with B"	12

1.0 INTRODUCTION

Long-term environmental assessments -- those covering one hundred years or more -- typically use results of computer simulations as a basis for their conclusions. The credibility of this approach depends on how well the models and data used in the simulations are accepted by reviewers. For acceptance, it is important that the scenarios simulated be plausible. One requirement for plausibility is that values of parameters in the simulation model display the same characteristics and relationships as do values observed in the field, laboratory, and literature. In particular, observations that are known to be correlated in the field should correspond to parameter values that are correlated in the model.

Correlation is defined to be the normalized expected value of the product of two random variables, denoted by C(X,Y):

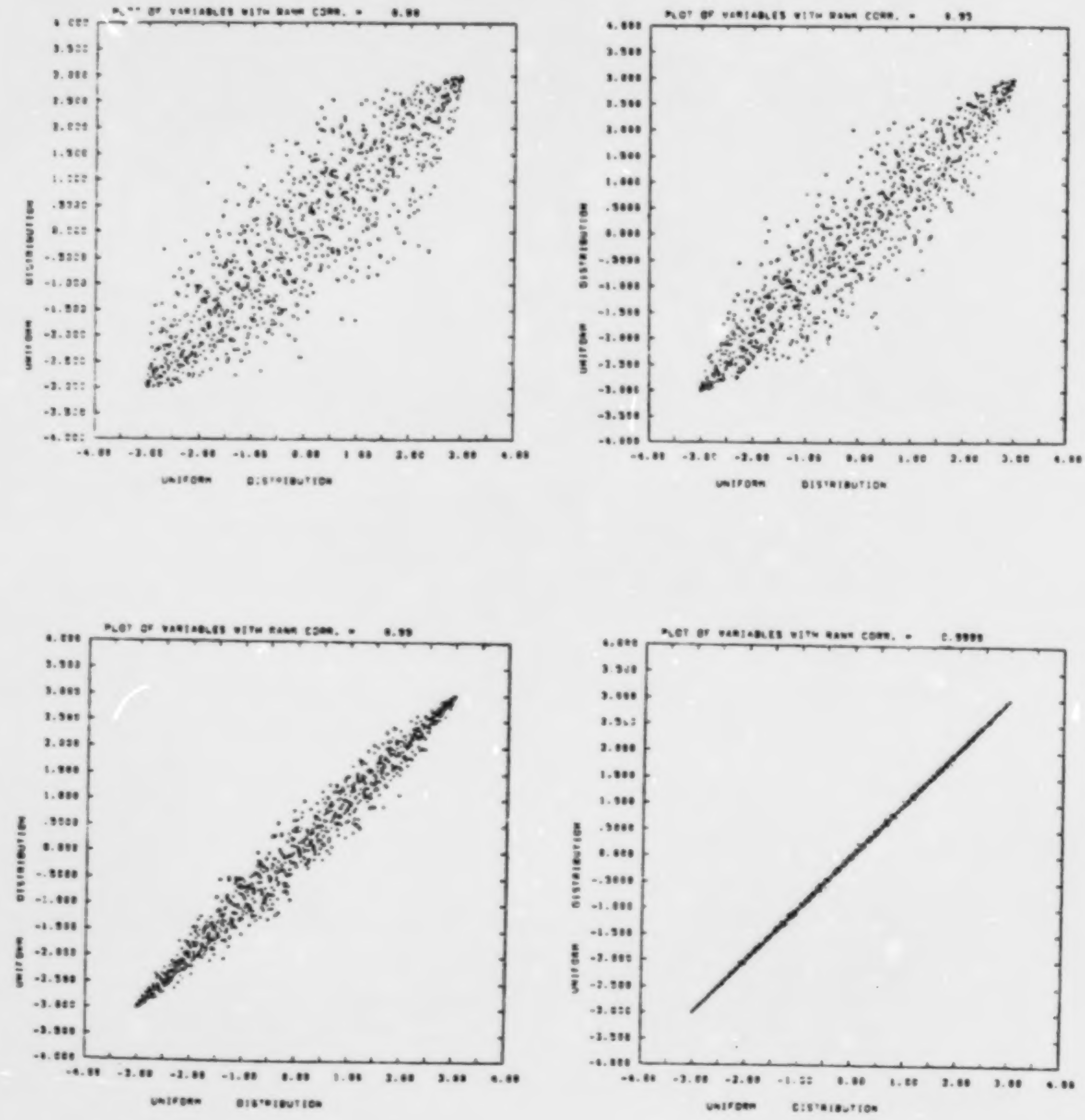
$$C(X,Y) = E\ (X-\mu_X)(Y-\mu_Y)\ /\sigma_X\ \sigma_Y \qquad (1)$$

Here X and Y are two random variables with means μ_X and μ_Y and standard deviations σ_X and σ_Y respectively. If X and Y are linearly related by a line with a positive slope, then $C(X,Y)=1$; if they are linearly related by a line with a negative slope, then $C(X,Y)=-1$; if they are independent, then $C(X,Y)=0$. Other values between -1 and 1 indicate varying degrees of linear dependence (see Figure 1).

Correlations can be observed in experimental data where two or more variables are measured under the same conditions. When variables are observed independently, correlation coefficients cannot be measured directly. For example, data gleaned from independent sources in the literature will not be correlated. In such cases, correlations can sometimes be hypothesized (but not

FIGURE I

SCATTERPLOTS SHOWING DIFFERENT LEVELS OF CORRELATION



SOURCE: ADAPTED FROM IMAN AND DAVENPORT, 1980 necessarily quantified) by experts having substantial experience with the natural system being modelled. In identifying a correlation, an expert is answering a question like: "If variable X is varied, how do I expect variable Y to respond, if at all?"

Correlations mark relationships in a natural system. Modelling of a natural system means finding ways to capture the major relationships in the system, as revealed by observational data and expert knowledge. Some relationships are modelled by explicit functions between variables. Explicit forms require that some variables be treated as more basic than others; the set of basic variables (i.e. the set of parameters) is used to calculate the rest. The designation of a basic or controlling variable is a human artifice imposed on the system model to render it tractable. Some relationships hopefully minor, are perforce left out, since it is difficult to model explicitly all the interactions in a natural system containing complex nets of feedback relationships. Residual effects appear as correlations among the parameters.

Parameter correlations can be input into the random sampling stage of a stochastic simulation model. If a mechanism for generating correlated parameters is available, known real-world correlations are then translated into sample correlations among the parameters.

That is not to say that all pairs of parameters in an environmental model can or should be explicitly correlated. For N parameters there are N(N-1)/2 distinct correlation coefficients; trying to estimate all of them accurately would not just strain current knowledge, but would require manyfold expansions of current research programs. Nevertheless, when data is available or can be estimated, selected correlation coefficients should be considered as prime candidates for inclusion in a probabilistic assessment.

The paper is intended to provide an overview of how correlated parameters may be treated in stochastic simulations, with particular emphasis on the National Uranium Tailings Program. Examples are in some cases drawn from the Canadian Nuclear Fuel Waste Management Program, with which the authors have considerable experience.

Section 2 gives an illustrative example of the use of correlation coefficients. Section 3 expands on the benefits to the assessment analyst in observing correlations. Section 4 describes techniques that can be used in determining correlations. Section 5 describes techniques that can be used to enforce correlations in sampling scenarios for stochastic simulations.

2.0 A CORRELATION EXAMPLE

Consider the following simplistic model for the dose consequences of a uranium tailings disposal system. There are three parameters:

X = Maximum Concentration of Radium-226 in a Lake
Y = Dose-Concentration Ratio for Radium-226 by Ingestion
Z = Dose-Concentration Ratio for Radium-226 by Inhalation

The model is:

$$D = X(Y + Z) \qquad (2)$$

where D is an estimate of maximum dose.

There are three correlation coefficients in this example:

C(X,Y) = Correlation Coefficient of X and Y
C(X,Z) = Correlation Coefficient of X and Z
C(Y,Z) = Correlation Coefficient of Y and Z

Since the order of variables in a correlation makes no difference (i.e. C(X,Y) = C(Y,X)), these are the only coefficients.

Suppose that experts have suggested that C(X,Y) = C(X,Z) = 0, and that C(Y,Z) has a significant value, say 0.7. That would mean that the two dose-concentration ratios would tend to be high together, and low together. What effect would the value of C(Y,Z) have on the value of dose D?

In this case the expected or mean dose, E (D), would be the same no matter what the value of C(Y,Z), since the two factors X and (Y + Z) are statistically independent (i.e. C(X,Y) = C(X,Z) = 0):

$$E(D) = E(X)\ E(Y) + E(Z) \qquad (3)$$

However, the variance of D would depend on C(Y,Z), since the more that Y and Z are positively correlated, the more they would tend to take on high values together and low values together, leading to more high doses and more low doses. On the other hand, a negative correlation between Y and Z would lead to compensating effects (Y high and Z low, or Y low and Z high) and a lower variance of D.

Consequently the results of an assessment based on this simplistic model would depend on the value of C(Y,Z) just as it depends on distributions for X, Y, and Z. There might even be some justification for sampling C(Y,Z) from a distribution to represent uncertainty in the expert's knowledge of its value. In effect, C(Y,Z) becomes a fourth parameter of the model.

If new evidence arises that suggests the dose-concentration ratios depend strongly on concentration, then C(X,Y) and C(X,Z) can no longer be set to zero. Both the mean and the variance of D depend on C(X,Y) and C(X,Z), so ignoring the new information would definitely skew the results of an assessment.

This effect is similar to that of ignoring any processes or parameters that are known to affect dose. The effects may or may not be large enough to cause concern. It is the challenge of the assessment analyst to include all the processes and parameters (including correlations) that significantly affect his results, while striving to keep his models and data simple enough for review by his peers.

3.0 REASONS FOR CORRELATING PARAMETER VALUES

3.1 General

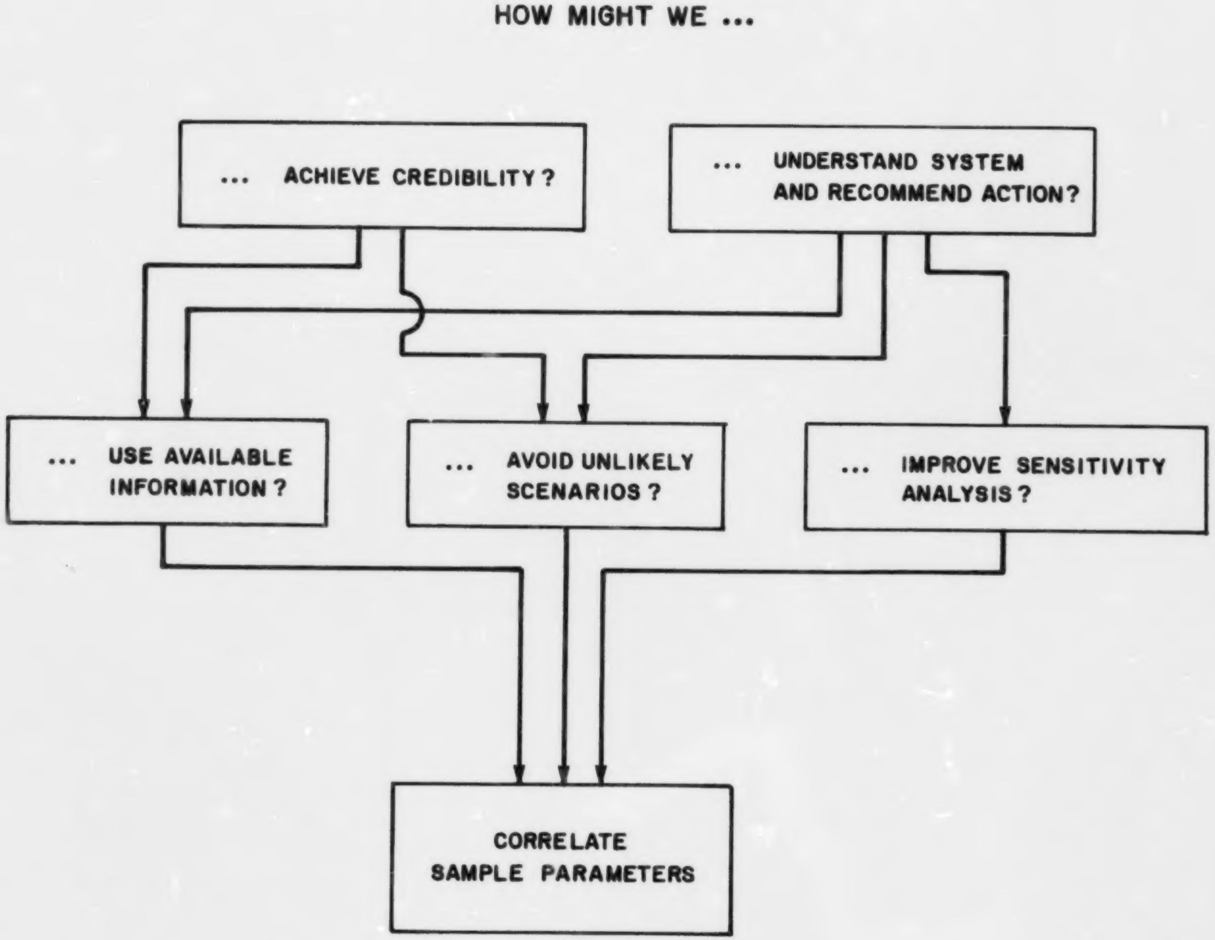
A simulation model used in environmental assessment should achieve credibility in the scientific community. It should also help assessment analysts understand the modelled system, thereby forming a sound basis for recommended action. The appropriate use of correlated variables in a simulation model helps it to achieve these goals in three ways (see Figure 2) as discussed in the following sections.

3.2 Use of Available Scientific Data and Expert Opinion

Scientists, on the basis of their experience, know that some real world observations are correlated. For example, limnologists (researchers in sedimentation) may know that the speed of sedimentation is similar for suspended solids and biomass. In the uranium tailings model, they may recommend that "suspended solids sedimentation velocity" and the "biomass sedimentation

FIGURE 2

REASONS FOR CORRELATING PARAMETER VALUES



HOW MIGHT WE ...

... ACHIEVE CREDIBILITY?

... UNDERSTAND SYSTEM AND RECOMMEND ACTION?

... USE AVAILABLE INFORMATION?

... AVOID UNLIKELY SCENARIOS?

... IMPROVE SENSITIVITY ANALYSIS?

CORRELATE SAMPLE PARAMETERS

velocity" be correlated. Similarly, scientists may recommend that the "radium-226 bioaccumulation factor by aquatic biota" be correlated with the "radium-226 transfer coefficient: water to fish flesh". A simulation model that captures and uses this information from the start will have greater credibility and lead to better understanding and better recommendations than one that doesn't.

3.3 Avoiding Unlikely Simulation Scenarios

A scientist reviewing simulation results may be disconcerted to find scenarios that seem unusual, or even impossible. For example, a scenario may have slow leaching of radium-226 from tailings combined with fast oxidation of the pyritic material in the tailings mass. If the conditions that lead to slow leaching are incompatible with fast oxidation, then the scenario may be untenable in a way that leads to over-optimistic results.

A scientist who wants to criticize the foundations of the assessment may select instances where seemingly contradictory values have been assigned to related parameters. When presented to regulators, decision-makers, or the public, these instances may undermine the credibility of the model, and they may flag skewed recommendations based on the model. If related parameters have been correlated in sampling, scenarios will be more plausible, and these problems will be less likely to occur.

3.4 Improved Sensitivity Analysis

Recommendations for site selection or design criteria may be founded on a model-based understanding of which parameters strongly affect simulation results. When parameters are correlated, indirect paths arise for parameters to influence simulation results. For example, parameter A may have little effect on a particular consequence, say radon dose, but it may be strongly correlated with parameter B that does affect radon dose. Then in site selection, parameter A may be used as a selection criterion because it serves as a marker for parameter B. Therefore by using correlations in sampling parameters, better understanding of parameter influences and better recommendations may be obtained.

4.0 METHODS OF ASCERTAINING PARAMETER CORRELATIONS

4.1 General

Given that including known correlations between model parameters is required for a long-term assessment, methods are needed for ascertaining correlations. The problems in developing such methods are partly scientific (how to measure correlations in the field) and partly psychological (how to question experts to obtain consistent and unbiased information about correlations expected).

4.2 Research Program

The scientific questions can be tackled by a research program. There are applications for straightforward field observations and for careful applications of the scientific method. For example, a researcher may conjecture that the oxygen diffusivity through the tailings varies inversely with the tailings bulk density. This hypothesis can be tested with a set of experiments. Of particular interest are hypotheses that identify "controlling" or "basic" variables, upon which all the others appear to depend. If supported by the evidence, such hypotheses enhance our understanding of the environment being modelled, and simplify the models that can be used.

One goal in selecting basic parameters is to choose a set without strong correlations, because strong correlations among the basic parameters suggest significant relationships in the real system that have not been explicitly modelled. Where this goal is not attainable, a lesser goal would be to measure explicitly the correlations that occur.

4.3 Eliciting Expert Opinion

Where correlations are a matter of theory or expert opinion, rather than experimental data, care must be taken in eliciting this information. A scientist may not be familiar with the concept of correlation, and so must be guided in expressing his theoretical understanding or cultivated opinion. Graphic representations of the relationship between two variables, as in Figure 1, may help. Alternatively, the expert may be able to define a mathematical model of the relationship as he understands it. This model may be incorporated as part of the environmental model, or it may be used in a separate simulation to estimate correlation coefficients.

Where a group of experts are to estimate correlations, group decision-making techniques such as the Delphi procedure (Linstone and Truoff 1975) may be of assistance. In the Delphi approach, the experts individually and anonymously express their opinions based on theory or experience. Then in several rounds of revision, each expert gets a chance to review the contributions of the others, and change his own written opinion. Often this approach converges to one answer.

A number of studies have been devoted to strategies for eliciting expert opinion, either individually or in groups. Much of this literature is reviewed in Stillwell et al (1982).

4.4 Multivariate Correlations

The most challenging situation is one where several parameters are mutually dependent. Then a matrix of correlation coefficients or functional relationships is required. In a very few cases such a matrix may emerge from a research program, but in most cases it will need to be estimated on the basis of imperfect knowledge. It is difficult to construct a valid correlation matrix to represent expert opinion. A correlation matrix is a symmetric matrix with 1's on the diagonal and numbers between -1 and 1 everywhere else, but not all such matrices are valid correlation matrices.

Using the simple model from Section 2, one could assume that $C(X,Y) = C(X,Z) = 1$, but $C(Y,Z) = -1$. This set of assumptions is nonsensical, for it can be restated as the three statements:

S1. As X increases, Y increases linearly with X.
S2. As X increases, Z increases linearly with X.
S3. As Y increases, Z decreases linearly with Y.

Here S3 contradicts S1 and S2, which together imply that Z would increase linearly with Y. With more variables, it would be more difficult to detect inconsistencies by inspection.

Mathematically, a correlation matrix must have the properties mentioned above and be positive semidefinite (have all non-negative eigenvalues) to be valid. This requirement constrains what an expert may specify for correlations, or what a random sampling routine could generate if correlation coefficients were sampled randomly.

To generate a valid correlation matrix to represent expert opinion (or even sample on randomly), at least two approaches can be used. The first approach (used in the IMSL routine GGCOR (IMSL 1984) is to start with a set of positive eigenvalues, generate an orthogonal matrix from a set of independent basis vectors, and then form the correlation matrix using similarity transformations. The IMSL routine generates random correlation matrices; some research would be required to use this technique to capture expert opinion, or to generate correlation matrices with specific properties.

The second approach is simpler but less flexible. One parameter is chosen to be sampled independently. The remaining parameters are sampled using the Kleijnen technique (or any other) to be correlated with the first. They will also be correlated among themselves by virtue of being correlated to the same parameter, but no effort is made to control their mutual correlation coefficients. If any of these correlation coefficients can be estimated accurately, this method will fail because it does not represent available information.

For example, using the second method on the parameters from the example in Section 2, only two out of the three correlation coefficients are explicitly determined. If Z is chosen to be sampled independently, then C(X,Z) and C(Y,Z) must be supplied, and C(X,Y) is determined as a result of the sampling technique. C(X,Z) and C(Y,Z) are unconstrained. They may be assigned any values between -1 and 1; hence expert opinion may be solicited or these values may be sampled from distributions. C(X,Y) will be determined accordingly, but not uniquely. The value of C(X,Y) obtained will depend on the technique used to sample values for X, Y, and Z.

5.0 METHODS OF GENERATING CORRELATED PARAMETER VALUES

5.1 General

Once correlations between parameters have been identified, they must be applied to the procedure that generates scenarios to be simulated. This step can be accomplished in a number of ways. One approach is to convert correlations into functional relationships, and to calculate the values of parameters that were once sampled. Another approach is to use random number generation techniques that enforce correlations. Kleijnen (1974) has proposed such a technique for Monte Carlo sampling. It preserves the correlation coefficient, the distribution of one of the parameters, and the mean and variance of the other, but may not preserve the shape of the second parameter's distribution. A generalization of this technique, as implemented in the IMSL routine GGNSM (IMSL 1984), can be used to sample parameters from a multivariate normal distribution. Iman and Conover (1981) have proposed another technique that yields approximately correct rank correlations. It has been widely applied to latin hypercube sampling, and also works with unstratified Monte Carlo sampling.

5.2 Constructing Functional Relationships

Where the form of a relationship between parameters is known or suspected, a functional relationship may be the easiest way to express the dependence. A functional relationship is a formula that expresses one parameter as a function of others. For example, the formula may be:

$$\begin{pmatrix} \text{suspended solids} \\ \text{sedimentation velocity} \end{pmatrix} = \begin{pmatrix} 8*\text{biomass} \\ \text{sedimentation velocity} \end{pmatrix}$$

In this example, "biomass sedimentation velocity" has been selected as more "basic" than "suspended solids sedimentation velocity". In defining a set of basic parameters, some decisions have to be made. The selection of basic parameters can be especially difficult in modelling natural systems, because it may not be obvious which variables, if any, are controlling a natural situation. If A and B are two correlated parameters, it may be appropriate to express A=F(B), or B=G(A), or perhaps A=F(C) and B=G(C) where C is a third parameter (see Figure 3). Such decisions are usually based on expert understanding of the system being modelled, whether theoretical or experimental.

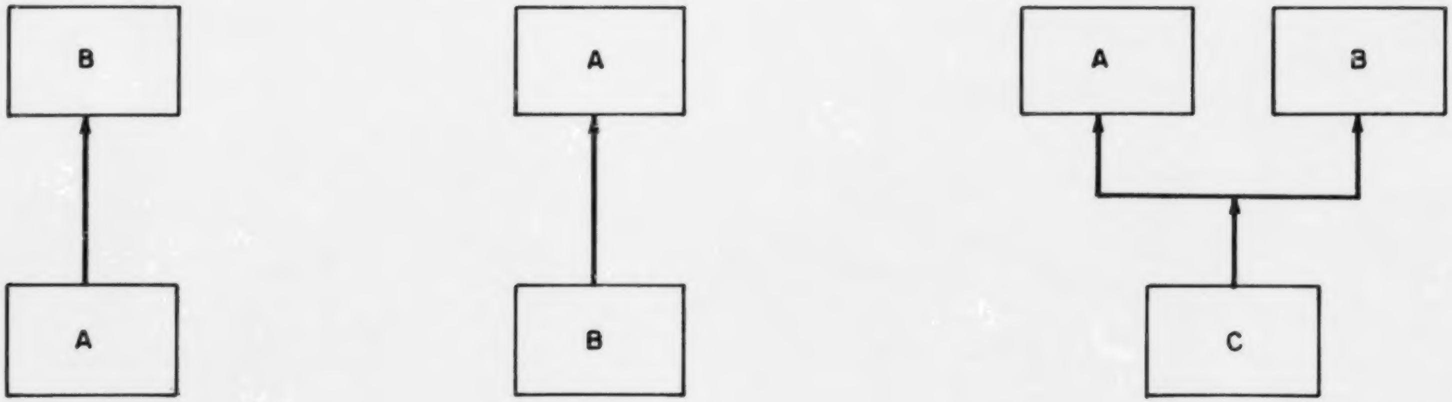
Often a formula introduces new constants, such as the constant "8" in the first example above. Such constants may have physical units, needed to make the formula dimensionally correct. In some cases these constants may be considered to be new parameters, with an appropriate distribution to be estimated. Treating them as parameters introduces variability into the relationship, while keeping the appropriate functional form. Therefore, the number of parameters (coefficients) used in expressing relationships tends to increase as more parameters are evaluated by functional relationships.

At the same time the number of sampled parameters having an easily understood physical meaning tend to go down as more functional relationships are used. These are the parameters that are observed in field experiments, and for which correlations are known or suspected. As a model evolves, some are chosen to be more basic than others, and the rest are expressed in terms of the basic parameters.

One of the benefits of developing functional relationships among parameters is the opportunity provided to put components of the model to experimental test. It is much easier to generate evidence to reject a faulty relationship than it is to reject a

FIGURE 3

POSSIBLE MEANINGS OF
"A IS CORRELATED WITH B"



correlation coefficient. A model that has been stringently tested and survives will draw more credibility than one that is based on theoretical considerations alone.

As an example of an evolving model, consider the model of uranium fuel solubility used in the Nuclear Fuel Waste Management Program. At first the solubility of uranium fuel was sampled as an independent parameter with a uranium solubility distribution defined using expert opinion (Wuschke et al 1981).

With greater understanding came the realization that there were two main sources of uncertainty represented by the solubility distribution: 1) experimental error in the thermodynamic data base used, and 2) uncertainty about chemical conditions (pH, Eh, etc.) in a disposal vault. The second source of uncertainty was by far the greater. This conclusion led to a more sophisticated analysis of the variability of uranium solubility in natural groundwater (Goodwin et al 1982).

Investigation of chemical conditions in the disposal vault revealed that these chemical correlations have a major effect on parameters other than uranium solubility. Many parameters (more than 60) in the Nuclear Fuel Waste Management model, such as the rate of corrosion of a disposal container, or the sorption of uranium, can be described using functional relationships which depend on a common set of basic sampled chemical parameters such as pH and Eh. By incorporating all these functional relationships into the model, correlations can be automatically introduced among the parameters that depend on chemical conditions. Additional variability can be introduced into the model as required by giving distributions to constants that appear in the formulas used. This work is continuing.

5.3 Kleijnen's Correlated Generation Technique

The technique described by Kleijnen (1974) for generating correlated variates is quite straightforward. Assume X and Y have means μ_X and μ_Y and standard deviations σ_X and σ_Y respectively. Let Z be a normal variate with mean 0 and standard deviation 1. Then to sample both X and Y in a way that preserves the distribution of X, the mean and variance of Y, and the correlation coefficient C(X,Y) use the following algorithm:

1. Sample X and Z from their distributions.
2. Evaluate $Y = aX + b + u$
   where $a = C(X,Y)\ \sigma_Y / \sigma_X$

$$b = \mu_Y - C(X,Y)\ \mu_X\ \sigma_Y / \sigma_X$$

$$u = \sigma_Y\ Z(1-C(X,Y)^2)^{1/2}$$

For example, suppose X and Y are both normally distributed, X having mean 5 and standard deviation 1, Y having mean 100 and standard deviation 20. Suppose the correlation coefficient C(X,Y) = 0.80. Then:

$$a = 16$$

$$b = 20$$

and $u = 12*Z$

Suppose X is sampled at 3.5 and the intermediate variable Z (normal with mean 0 and variance 1) takes the sampled value 0.5. Then:

$$Y = 16*X + 20 + 12*Z = 82$$

As expected, since X has a low value, Y also has a low value.

This algorithm preserves the distribution of Y as well as the distribution of X if both X and Y are normally-distributed. Otherwise, only the first two moments (mean and variance) of Y's distribution are correct, and its shape may vary from the given distribution. For example, if X was uniformly distributed between 3 and 7, and Y was distributed as above, then the same technique as above could be used, except that the standard deviation of X is now 16/12 = 1.33 instead of 1. The difference between the two cases would not be visible in any particular pair of values (X,Y) and an analysis of a large number of samples would reveal that neither X nor Y was normally distributed. X would be uniformly distributed as specified, and Y would have an unknown distribution, similar to a normal distribution, and with the mean and variance as originally specified.

Kleijnen's technique encounters some difficulty when the distributions involved are truncated. Small truncations (eg. at the 0.001 and 0.999 quantiles) should have little effect, but large truncations modify the sample correlation. Furthermore, the truncation scheme must be implemented carefully. While it is easy to truncate X, since it is generated first, it is more difficult to ensure that values for Y can be generated up to the truncation point, but no further. The problem of truncation has been solved by one of the authors (Andres, unpublished work), but the problem of its effect on the sample correlation coefficient remains.

Techniques for generating variables from a multivariate normal distribution are also available. For example, the IMSL routine GGNSM performs this function for normal variates with zero mean and a given covariance matrix. Users must rescale the results for non-zero means.

To date in the Nuclear Fuel Waste Management Program, insufficient data has been available to require this generality. Kleijnen's technique as described above has sufficed.

5.4 Rank Correlations by Iman and Conover's Technique

Rank correlations are found by taking correlations in a sample set where every parameter value has been replaced by its rank. Generally the rank of the largest value for a parameter is 1; that of the smallest is N, where N is the sample size. For highly skewed distributions (eg. lognormal), rank correlations have better behaved properties than correlations of the parameters themselves (Iman and Davenport 1980).

Iman and Conover (1981) have proposed a method for generating data sets where the parameters have approximately the required rank correlation matrix. It has been widely used with latin hypercube sampling (eg. Iman and Davenport 1980), and applies to unstratified Monte Carlo sampling as well.

The procedure has three essential steps:

1. Generate a sample set of N values for each of K parameters, and arrange them in an NxK matrix M, with one row for each simulation, and one column for each parameter.
2. Generate an NxK matrix R with the required rank correlations among its columns.
3. Copy the rank correlation structure from R to M by permuting the entries in each column of M. When finished, the rank order in each column of M should be the same as the rank order of the corresponding column of R.

This technique has the advantage that it preserves the originally-sampled values, whatever their distribution. That is the reason that it works with any distribution and with both latin hypercube and unstratified Monte Carlo samples.

How well the desired rank correlation matrix is reproduced in M depends on the algorithm in step 2. The algorithm used by Iman and Conover is only approximate: they actually generate a matrix R with a correlation matrix the same as the desired rank correlation matrix, and then assume that the rank correlation matrix of R will be similar to its correlation matrix.

Even if all parameters in a simulation are thought to be independent, this algorithm can be of use in enforcing a rank correlation matrix that is the KxK identity matrix. In this way, spurious correlations introduced in sampling can be reduced before the simulations take place.

Unlike the Kleijnen scheme, this technique encounters no difficulty with truncated distributions, nor the non-normal distributions. It is entirely distribution free.

The chief obstacle to the use of this technique is the size of the matrices involved. For N=1000 and K=30, matrices M and R are both size 1000x30. At this size, both the formation of R and the permutation of M can be time and space-consuming processes. For larger simulations (eg. N=2000 and K=500, as run in the Nuclear Fuel Waste Management Program), the cost of generating R and permuting M may be prohibitive.

6.0 CONCLUSION

The paper has reviewed some of the techniques required to treat correlated parameters in a stochastic simulation code. It has shown the benefits of treating correlated parameters appropriately in environmental assessments. Techniques for determining correlations, and applying correlations in sampling were also discussed.

The main point of this report is that correlations among sampled parameters in environmental models can and should be treated, where data is available. Existing methods for treating correlations, such as those mentioned in this paper, can be used in a straightforward way. They will improve both the credibility of environmental models, and the quality of recommendations based on them.

REFERENCES

Goodwin, B.W., LeMire, R.J., and Johnson, L.H., 1982. "A Stochastic Model for the Dissolution of Irradiated UO_2 Fuel." Proceedings of the International Conference on Radioactive Waste Management, September 12-15, 1982, Winnipeg, Manitoba.

Iman, R.L. and Conover, W.J., 1981. "A Distribution-free Approach to Inducing Rank Correlation Among Input Variables for Simulation Studies." Technometrics.

Iman, R.L. and Davenport, J.M., 1980. "Rank Correlation Plots for Use with Correlated Input Variables in Simulation Studies." Sandia National Laboratories Report SAND80-1903.

IMSL, 1984. "User's Manual -- IMSL Library, Volume 2."

Kleijnen, Jack P.C., 1974. "Statistical Techniques in Simulation - Part 1." Marcel Dekker, Inc., New York.

Linstone, H., and Turoff, M., 1975. "The Delphi Method: Techniques and Applications." Addison-Wesley, Reading, MA.

Stillwell, W.G., Seaver, D.A., and Schwartz, J.P., 1982. "Expert Estimation of Human Error Probabilities in Nuclear Power Plant Operations: A Review of Probability Assessment and Scaling." U.S. Nuclear Regulatory Commission Report NUREG/CR-2255.

Wuschke, D.M., Mehta, K.K., Dormuth, K.W., Andres, T.H., Sherman, G.GR., Rosinger, E.L.J., Goodwin, B.W., Reid, J.A.K. and Lyon R.B., 1981. "Environmental and Safety Assessment Studies for Nuclear Fuel Waste Management - Volume 3: Post-Closure Assessment." Atomic Energy of Canada Limited, Technical Record TR-127-3, Pinawa, Manitoba.

Discussion Paper on

MODELLING RADIUM EXCHANGE BETWEEN THE WATER COLUMN AND LAKE SEDIMENT

by
J.M. SCHARER AND B.E. HALBERT
SENES CONSULTANTS LIMITED
WILLOWDALE, ONTARIO
June, 1985

Abstract

The mechanisms affecting Ra-226 removal from the water column of a reference lake, exchange between the bottom sediment and the water column and burial in deep sediment are quantitatively assessed and described. Factors considered in the analysis include: biomass productivity; Ra-226 distribution between solid and liquid phases; sediment deposition, decomposition and burial, and Ra-226 diffusion out of the sediment layer. These factors are combined in the analysis to derive rate constants which are used in the mainstream modelling activities. Ranges of possible values for the rate constants are presented for use in probabilistic analysis.

TABLE OF CONTENTS

		Page
1.0	INTRODUCTION	1
2.0	CONCEPT OF REFERENCE LAKE SEDIMENT MODEL	2
3.0	SEDIMENTATION FROM THE LAKE WATER COLUMN	3
4.0	DISTRIBUTION COEFFICIENT OF SEDIMENTING MATERIAL (k_D^S)	4
5.0	DISTRIBUTION COEFFICIENT OF STABLE (DECAYED) SEDIMENT (k_D^D)	4
6.0	SEDIMENT DENSITY AND DEPOSITION RATE	5
7.0	DIFFUSION OUT OF THE SEDIMENT EXCHANGE LAYER	7
8.0	DERIVATION OF RATE CONSTANTS FOR RADIUM EXCHANGE	12
9.0	ESTIMATION OF VALUES FOR RATE CONSTANTS	14
	REFERENCES	17

1.0 INTRODUCTION

The movement of radionuclides through the aquatic pathways depends on numerous physical, biological and chemical processes which affect the resulting concentrations in water, sediment and aquatic biota. Since exposure to man may result from consumption of contaminated water or biota (eg. fish flesh), estimation of the radionuclide concentrations in these media is of particular interest in radiological assessments, especially in Canada where the aquatic pathways have generally been found to be the largest contributors to dose. Typically, the concentrations in water have been estimated taking into account only the effects of dilution and perhaps removal to bottom sediments. The concentration in fish flesh has usually been calculated using the aggregate, one-parameter transfer factor model.

Considering the importance of the aquatic pathways, one of the objectives set for the Phase II probabilistic assessment studies entailed investigating the adequacy of the aggregate modelling approach for predicting radium-226 (Ra-226) concentrations in the reference lake. This paper describes the details of the model developed to predict Ra-226 removal from the lake water column to the bottom sediment, release from the bottom sediment back to the water column, removal from the exchangeable sediment layer by burial, and dilution with inflow from the drainage basin. In a complementary paper, simple and multi-component models for estimating Ra-226 transfer to fish flesh are evaluated (see Gardner et al 1985).

2.0 CONCEPT OF REFERENCE LAKE SEDIMENT MODEL

Radium is introduced into the reference lake with the inflow from the upstream watershed. Removal of Ra-226 from the aqueous phase occurs through adsorbtion on inorganic solids and organic matter (including aquatic biota). Ultimately, a portion of the Ra-226 enters the lake sediments with settling solids. As a result of the oxidation of organic matter at the lake bottom, a portion of the Ra-226 is released from the sediment and re-enters the aqueous phase. The remaining fraction eventually becomes buried in the sediment and is thus unavailable for exchange with the water column.

A conceptual schematic of the lake sediment model is shown in Figure 1 and comprises two active compartments, the water column and the exchangeable sediment. The basic parameters defined for the assessment of Ra-226 transfer between the water column and sediment are given as follows:

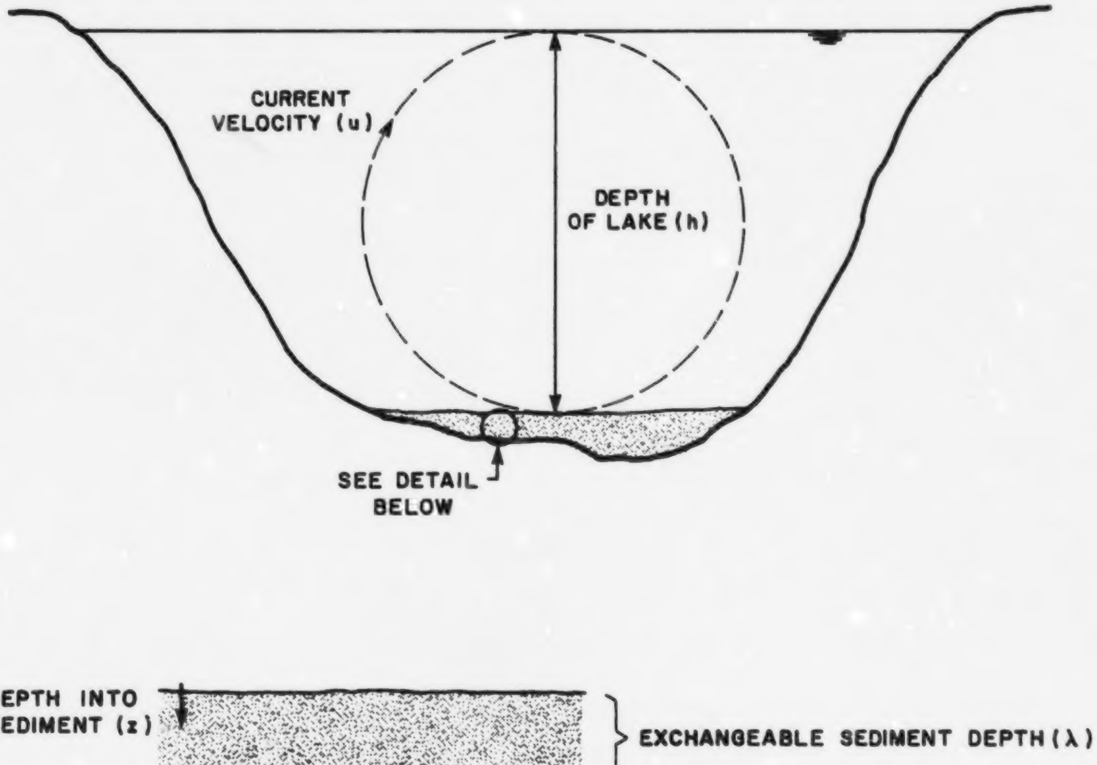
h = water column height (m)
λ = thickness of the sediment exchange zone (m)
C_w = concentration of Ra-226 in the water column ($Bq.m^{-3}$)
$\overline{C}$ = average interstitial water concentration of Ra-226 in the exchanging sediment ($Bq.m^{-3}$)
C_λ = concentration of Ra-226 at λ in the sediment interstitial water ($Bq.m^{-3}$)
u = current velocity in lake ($m.yr^{-1}$)

The accumulation of Ra-226 (as a flux term) in the exchangeable sediment (thickness $= \lambda$) is given by:

$$\lambda \frac{d\overline{C}}{dt} = \underbrace{\left[\sigma_S\, k_D^S\, C_w\right]}_{\substack{\text{deposition flux}\\ \text{of Ra-226 in}\\ \text{fresh sedimenting}\\ \text{matter}}} - \underbrace{\left[k_L(C_\lambda - C_w)\right]}_{\substack{\text{flux of Ra-226}\\ \text{from sediment}\\ \text{into the water}\\ \text{column}}} - \underbrace{\left[\sigma_D k_D^D\, C_\lambda\right]}_{\substack{\text{flux of Ra-226}\\ \text{out of the exchange}\\ \text{zone due to continuous}\\ \text{deposition of sediment}\\ \text{(i.e. loss due to burial)}}} \qquad (2.1)$$

FIGURE 1

CONCEPTUAL SCHEMATIC OF REFERENCE LAKE



where:

σ_S = rate of sedimenting material removal from the lake water column ($g.m^{-2}.yr^{-1}$)

k_D^S = average distribution coefficient for Ra-226 between sedimenting material (both organic and inorganic fractions) and the lake water column ($m^3.g^{-1}$)

k_L = mass transfer coefficient of Ra-226 into the lake water column from the sediments ($m.yr^{-1}$)

σ_D = rate of sediment burial (or removal from the active exchange layer)($g.m^{-2}.yr^{-1}$)

k_D^D = average distribution coefficient for Ra-226 in stabilized lake-bottom sediment ($m^3.g^{-1}$).

Collecting the lake water column concentration terms (C_w) in equation (2.1) yields:

$$\lambda \frac{d\overline{C}}{dt} = \left[\underbrace{(\sigma_S k_D^S + k_L)}_{\substack{\text{apparent net deposition} \\ \text{(settling and diffusion)}}} C_W \right] - \left[k_L \; C_\lambda \right] - \left[\sigma_D k_D^D \; C_\lambda \right] \quad \ldots . \quad (2.2)$$

3.0 SEDIMENTATION FROM THE LAKE WATER COLUMN

The lake system for the reference case is assumed to be an oligotrophic lake situated on the Canadian Precambrian Shield. Sedimentation is assumed to be characteristic of a low productivity lake environment. For a net primary production on a dry weight solids basis of 100 $g.m^{-2}.yr^{-1}$ (range 40 to 200 $g.m^{-2}.yr^{-1}$) (Wetzel 1975) and an organic content of 55% in the sedimenting solids, the sedimentation flux equals 180 $g.m^{-2}.yr^{-1}$.

4.0 DISTRIBUTION COEFFICIENTS OF SEDIMENTING MATERIAL (k_D^S)

The distribution coefficient for calculating Ra-226 removal from the water column on sedimenting matter has been found to differ between the organic and inorganic solids. Based on the results of lake sediment and biomass analyses reported elsewhere (SENES and SRC 1984, Swanson 1985) the distribution coefficients selected were:

.organic matter k_D: 12,000 $cm^3.g^{-1}$ dry solids = $12.0 x 10^{-3}$ $m^3.g^{-1}$ dry solids

.inorganic matter k_D: 3,500 $cm^3.g^{-1}$ dry solids = $3.5 x 10^{-3}$ $m^3.g^{-1}$ dry solids

The average distribution coefficient for sedimenting material therefore was calculated to be:

$$k_D^S = (0.55)(12.0 \times 10^{-3}) + (0.45)(3.5 \times 10^{-3})$$

$$k_D^S = 8.18 \times 10^{-3} \ m^3.g^{-1} \text{ dry solids}$$

thus,

$$\sigma_S k_D^S = 180 \ g.m^{-2}.yr^{-1} \times 8.18 \times 10^{-3} \ m^3.g^{-1}$$
$$= 1.47 \ m.yr^{-1}$$

5.0 DISTRIBUTION COEFFICIENT OF STABLE (DECAYED) SEDIMENTS (k_D^D)

The organic content of sedimented material undergoes continuous decomposition thus reducing the organic fraction. Typically, stabilized lake sediment contains approximately 23% organic matter (Swanson 1985) which is a sharp reduction from the 55% organic fraction characteristic of the fresh sedimenting material. The remaining 77% comprises the original 80 $g.m^{-2}.yr^{-1}$ of inorganic matter.

The flux of matter out of the exchange zone is equal to the flux of stable (decayed) sediments, i.e.:

$$80 \text{ g.m}^{-2}\text{.yr}^{-1} \text{ (inorganic)} \left[\frac{1}{1-0.23} \right]$$

$$= 104 \text{ g.m}^{-2}\text{.yr}^{-1}$$

Assuming that the average distribution coefficient for radium in the stable sediment may be calculated on the same basis as previously stated for fresh organic matter, taking into account the differences in the organic and inorganic fractions:

$$k_B = (0.23)(12.0\text{x}10^{-3}) + (0.77)\ (3.5 \text{ x } 10^{-3})$$

$$= 5.46 \text{ x } 10^{-3} \text{ m}^3\text{.g}^{-1}$$

thus,

$$\sigma_D\ k_B = (104)\ (5.46 \text{ x } 10^{-3})$$

$$= 0.568 \text{ m.yr}^{-1}$$

6.0 SEDIMENT DENSITY AND DEPOSITION RATE

The densities of the individual components of stable sediment were assumed to be:

$$\rho_{\text{inorganic}} = 2.45 \text{ g.cm}^{-3} \text{ or } 2.45 \text{ x } 10^6 \text{ g.m}^{-3}$$

$$\rho_{\text{organic}} = 1.05 \text{ g.cm}^{-3} \text{ or } 1.05 \text{ x } 10^6 \text{ g.m}^{-3}$$

On the basis of 1 g sediment having an organic content of 23%, the true density of stabilized sediment solids was determined as follows:

$$\rho_s = \left[\frac{1 \text{ g}}{\dfrac{0.23 \text{ g}}{1.05 \text{ g.cm}^{-3}} + \dfrac{0.77 \text{ g}}{2.45 \text{ g.cm}^{-3}}} \right] = 1.88 \text{ g.cm}^{-3} \text{ or } 1.88 \text{ x } 10^6 \text{ g.m}^{-3}$$

To calculate the volume fraction of solids (ε_s), a moisture content of 89% by weight was used based on the results of the field studies reported by Swanson (1985). For a moisture density of 1.0 $g.cm^{-3}$ and a solids density of 1.88 $g.cm^{-3}$, the volume fraction occupied by the solids equals:

$$\varepsilon_s = \left[\frac{\dfrac{0.11 \text{ g solid}}{1.88 \text{ g.cm}^{-3}}}{\dfrac{0.11 \text{ g solid}}{1.88 \text{ g.cm}^{-3}} + \dfrac{0.89 \text{ g } H_2O}{1.00 \text{ g.cm}^{-3}}} \right] = 0.06 \text{ cm}^3.\text{cm}^{-3}$$

Thus the porosity, volume fraction of liquid ε, equals:

$$\begin{aligned} \varepsilon &= 1-0.06 \\ &= 0.94 \text{ cm}^3.\text{cm}^{-3} \end{aligned}$$

The bulk density of sediment therefore may be calculated as follows:

$$\begin{aligned} \rho &= (1.88)\ (0.06) + 1.00\ (0.94) \\ &= 1.053 \text{ g.cm}^{-3} \end{aligned}$$

Similarly, the annual rate of sediment deposition may be calculated from:

$$\left[104 \text{ g.m}^{-2}.\text{yr}^{-1}\right] * \left[\frac{1 \text{ m}^3 \text{ (solid)}}{1.88 \times 10^6 \text{ g (solid)}}\right] * \left[\frac{1 \text{ m}^3 \text{ (bulk sediment)}}{0.06 \text{ m}^3 \text{ (solid)}}\right]$$

$$= 9 \times 10^{-4} \text{ m.yr}^{-1}$$

$$= 0.09 \text{ cm.yr}^{-1}$$

This value is comparable to net sediment deposition velocities reported in the literature for Canadian Shield lakes of 0.07 to

0.8 cm.yr^{-1} (Kipphut 1978). It also falls between the values reported by Kemp and Harper (1977) on Georgian Bay (0.03 cm.yr^{-1}) and Lake Huron (0.15 cm.yr^{-1}).

7.0 DIFFUSION OUT OF THE SEDIMENT EXCHANGE LAYER

To estimate the range of the mass transfer coefficient (k_L in equation (2.1)), two assumptions concerning the sedimenting material were made. The first assumption was that the organic content decays rapidly at the sediment surface before significant burial by fresh sedimenting material occurs. In this case, mass transfer is controlled by a laminar-flow film next to the sediment boundary. The estimation of the mass transfer coefficient requires the application of Fick's law for unsteady state diffusion:

$$\frac{\partial C}{\partial t} = - D \frac{\partial^2 C}{\partial z^2} \qquad (7.1)$$

Equation (7.1) has been solved by Higbie (1935) by applying his penetration theory for diffusion into a "semi-infinite" medium represented by the water column above a flat (sediment) surface as follows:

$$k_L = \sqrt{\frac{D}{\tau}} \qquad (7.2)$$

where

τ = the surface renewal time (yr)

D = diffusivity coefficient (m^2.yr^{-1})

The problem is the evaluation of the surface renewal time (τ). For a lake of mean depth "h", an approximate value for surface renewal may be taken as a function of the mean depth "h" (or depth of the hypolimnion during stratification) and mean current velocity "u":

$$\tau = \frac{\pi h}{u} \qquad (7.3)$$

In relation to the above schematic, the diffusion out of the sediment exchange layer may be equated as:

$$\left(\begin{matrix}\text{rate of diffusion}\\ \text{through sediment}\end{matrix}\right) = \left(\begin{matrix}\text{rate of transport}\\ \text{from surface of}\\ \text{sediment}\end{matrix}\right)$$

which translates to:

$$D \frac{d^2C}{dz^2} = \left(\frac{u}{\pi h}\right) C_W \qquad (7.4)$$

The approximate solution to equation (7.4) is given by:

$$\left.\frac{dC}{dz}\right|_{z=0} = \left(\frac{u}{\pi Dh}\right)^{1/2} (C_\lambda - C_w) \qquad (7.5)$$

or

$$\text{Flux} = D \left.\frac{dC}{dz}\right|_{z=0} = \left(\frac{uD}{\pi h}\right)^{1/2} (C_\lambda - C_w) = k_L(C_\lambda - C_w)$$

where $k_L = \left(\frac{uD}{\pi h}\right)^{1/2}$

Equation (7.5), known as the Dobbins-O'Connor relationship, has been used successfully to describe mass transfer in lakes and rivers (Baily and Ollis 1977). Typical values for the parameters in equation (7.5) are:

$D = 6.36 \times 10^{-6}\ cm^2.s^{-1} = 0.02\ m^2.yr^{-1}$

$u = 0.1$ to $1.0\ cm.s^{-1} = 3.15 \times 10^4$ to $3.15 \times 10^5\ m.yr^{-1}$

Mean depths of 5m and 10m were selected to assess the effect of lake depth (h) on the mass transfer coefficient (k_L). Using extreme values for the parameters, k_L values were calculated as follows.

Case 1:
If h = 5m and $u = 0.1\ cm.s^{-1}$ then
$k_L = 6.33\ m.yr^{-1}$

Case 2:
If h = 5m and $u = 1.0\ cm.s^{-1}$ then
$k_L = 20.0\ m.yr^{-1}$

Case 3:
If h = 10m and $u = 0.1\ cm.s^{-1}$ then
$k_L = 4.48\ m.yr^{-1}$

Case 4:
If h = 10m and $u = 1.0\ cm.s^{-1}$ then
$k_L = 14.2\ m.yr^{-1}$

Hence, the k_L value is seen to range from about 4.5 to 20 $m.yr^{-1}$.

The second assumption involves a slow decay of the organic matter. In this case the sediment would be buried before complete decomposition could take place. Since this is used to calculate the minimum value of the radium-226 flux back into the water column, the mixing of the exchange layer by bioturbation is assumed to be negligible. The decay is equivalent to a "reaction rate". The radium-226 content of the porewater in this case can be represented by the following differential equation:

$$\frac{\partial C}{\partial t} + D\,\frac{\partial^2 C}{\partial z^2} - k_t C = 0 \qquad (7.6)$$

where

k_t = time constant (yr^{-1})

C = concentration of dissolved Ra-226 in the porewater ($Bq.m^{-3}$)

If the decomposition continues as long as the sedimented material remains in the exchange zone ($z < \lambda$), the instantaneous value of the time constant is given by:

$$k_t = k \left[\frac{(\sigma_S k_D^S - \sigma_D k_D^D)\, e^{-kt}}{\sigma_S k_D^S + (\sigma_S k_D^S - \sigma_D k_D^D)\, e^{-kt}} \right] \qquad (7.7)$$

where

k = the first order decay constant of the adsorbed radium flux.

In equation (7.7), the decay of organic matter, hence, the decrease in Ra-226 adsorbtion capacity is assumed to be exponential. As the adsorbtion capacity of the sediment declines, the desorbed radium is transported back into the water column. If the sedimentation velocity (u) is assumed constant, the time period (t) in equation (7.7) can be replaced by $t=z/u$ and the steady state diffusion equation (7.6) becomes the following:

$$D\,\frac{d^2C}{dz^2} = k \left[\frac{(\sigma_S k_D^S - \sigma_D k_D^D)\, e^{-kz/u}}{\sigma_S k_D^S + (\sigma_S k_D^S - \sigma_D k_D^D)\, e^{-kz/u}} \right] C \qquad (7.8)$$

According to equation (7.8), the radium-226 sorption capacity, and hence the flux, varies with depth from $\sigma_S k_D^S$ at the surface of the sediment ($z=0$) to $\sigma_D k_D^D$ at the lower limit of the exchange zone ($z = \lambda$). Equation (7.8) can be numerically integrated after substitution of previously calculated parameter values ($\sigma_S k_D^S$, $\sigma_D k_D^D$, D, and u) and assuming a range of values for the first order rate constant (k). Using the flux expression given by equation (2.1), the calculated values of the mass transfer coefficient (k_L) at various ratios of C_λ/C_w are given in the following examples.

C_λ/C_w	k_L ($m.yr^{-1}$)
1.02	7.95
1.05	5.09
1.10	3.53
1.20	2.63
1.30	2.19
1.40	1.94
1.5	1.77
2.0	0.35

Since the C_λ/C_w ratio is not likely to exceed 2.0, the minimum value of k_L is taken as 0.35 $m.yr^{-1}$.

On the basis of the above two example analyses, the value of the mass transfer coefficient, k_L, is expected to range from 0.35 to 20 $m.yr^{-1}$. If the distribution is lognormal, the mean value of k_L equals 2.65 $m.yr^{-1}$.

Substitution of numerical values into equation (2.2) yields:

$$\lambda \frac{d\bar{C}}{dt} = (1.47 + 2.65)\, C_W - 2.65\, C_\lambda - 0.57\, C_\lambda \qquad (7.9)$$

$$= 4.12\, C_W - 3.22\, C_\lambda$$

Note that at steady state ($d\bar{C}/dt = 0$)

$$\frac{C_\lambda}{C_w} = \frac{4.12}{3.22} = 1.28$$

This analysis indicates that there is not a substantial difference between Ra-226 in the water column and the sediment porewater (or interstitial water).

8.0 DERIVATION OF RATE CONSTANTS FOR RADIUM EXCHANGE

The rate of change of the mass of Ra-226 in the water column and sediments can be expressed as follows:

$$\frac{d\{Ra_w\}}{dt} = k_2\{\overline{Ra}\}_s - k_1\{Ra\}_w \qquad (8.1)$$

and

$$\frac{d\{Ra_s\}}{dt} = k_1\{Ra\}_w - k_2\{\overline{Ra}\}_s - k_3\{\overline{Ra}\}_s \qquad (8.2)$$

where

$\{Ra\}_w$ = the mass of Ra-226 in the water column (Bq)

$\{\overline{Ra}\}_s$ = the average mass of Ra-226 in the sediment exchange zone (Bq)

k_1 = the rate of Ra-226 removal from the water column (yr^{-1})

k_2 = the rate of Ra-226 introduction back to the water column from the sediment (yr^{-1})

k_3 = the rate of loss of Ra-226 to deep sediment, i.e. the rate of burial (yr^{-1})

By definition, the mass of Ra-226 in the sediment is given by:

$$\{\overline{Ra}\}_s = A\lambda\ [k_D^D\ \rho_s\ (1-\varepsilon) + \varepsilon]\ \overline{C} \qquad (8.3)$$

where A is the surface area of the lake bottom, ρ_s, is the solids density and is the porosity. In the water column, the mass of Ra-226 equals:

$$\{Ra\}_w = AhC_w \qquad (8.4)$$

In equation (8.3), $\bar{C}$ is unknown, but our best estimate of $\bar{C}$ is:

$$\bar{C} = \frac{C_w + C_\lambda}{2} = \frac{\frac{C_\lambda}{1.28} + C_\lambda}{2} = 0.89\ C_\lambda \qquad (8.5)$$

Rearranging equation (8.5) we get:

$$C_\lambda = \frac{\bar{C}}{0.89} = 1.12\ \bar{C}$$

Substitution into equation (2.1) yields:

$$\lambda \frac{d\bar{C}}{dt} = \sigma_S\ k_D^S\ C_W - k_L\ (1.12\bar{C} - C_W) - \sigma_D k_D^D\ 1.12\ \bar{C} \qquad (8.6)$$

Multiplying equation (8.6) by the area A and re-ordering we get:

$$\lambda A \frac{d\bar{C}}{dt} = (\sigma_S k_D^S + k_L)AC_W - (1.12k_L)A\bar{C} - (1.12\ \sigma_D k_D^D)A\bar{C} \qquad (8.7)$$

Comparing equation (8.7) to equation (8.2), it is seen that:

$$\lambda A\ \frac{d\bar{C}}{dt} = d\ \frac{\{RA\}_S}{dt}$$

Thus collecting the terms for $\{Ra\}_W$ and $\{Ra\}_S$ yields:

$$k_1 = \frac{\sigma_S k_D^S + k_L}{h} \qquad (8.8)$$

$$k_2 = \frac{1.12\ k_L}{\lambda\ [k_D^D\ \rho_s\ (1-\varepsilon)\ +\varepsilon]} \qquad (8.9)$$

$$k_3 = \frac{1.12\ \sigma_D\ k_D^D}{\lambda\ [k_D^D\ \rho_s (1-\varepsilon)\ +\varepsilon\ ]} \qquad (8.10)$$

9.0 ESTIMATION OF VALUES FOR RATE CONSTANTS

The rates of radium exchange between the water column and the sediment (given by k_1 and k_2) and the rate of permanent loss to the deep sediment (given by k_3) can be determined by substitution of appropriate values for the independent parameters in equations (8.8), (8.9) and (8.10). Nominal values for the rate constants were calculated by substituting the typical values estimated previously for several of the independent parameters, namely:

$\sigma_s k_D^S = 1.47\ \text{m.yr}^{-1}$
$\sigma_D k_D^D = 0.57\ \text{m.yr}^{-1}$
$k_L = 2.65\ \text{m.yr}^{-1}$
$k_D^D = 5.46 \times 10^{-3}\ \text{m}^3.\text{g}^{-1}$
$\rho_s = 1.88 \times 10^6\ \text{g.m}^{-3}$
$\varepsilon = 0.94\ \text{m}^3.\text{m}^{-3}$

In addition, the following values were selected for the remaining unspecified parameters:

$h = 10$ m (ie. equal to the mean depth selected for the reference lake)

$\lambda = 0.05$ m (i.e. the assumed effective mixing depth in the lake sediment due to bioturbation by benthic invertebrates)

The calculated nominal values obtained for the rate constants were:

$k_1 = 4.1 \times 10^{-1}\ yr^{-1}$
$k_2 = 9.6 \times 10^{-2}\ yr^{-1}$
$k_3 = 2.1 \times 10^{-2}\ yr^{-1}$

The rate of radium release back to the water column from the bottom sediment (k_2) is seen to be significant, i.e. equal to approximately 25% of the rate of removal from the water column (k_1). The rate of radium loss to the deep sediment (k_3) in contrast is seen to equal a minor amount, i.e. equal to about 5% of the rate of radium removal from the water column. This latter observation is due mainly to the slow rate of sediment accumulation on the bottom of oligotrophic lakes situated on the Canadian Shield.

The incorporation of the rate constant relationships given by equations (8.8), (8.9), and (8.10) into a probabilistic analysis requires that the uncertainties in the input parameters be accounted for. The rate constants may be calculated in the probabilistic analysis by substituting randomly selected values for the independent parameters in equations (8.8), (8.9) and (8.10).

Alternatively, the range of possible values for the rate constants (k_1, k_2 and k_3) can be assessed beforehand and distributions for the rate constants can be specified as input to the probabilistic analysis. This approach was taken. Minimum and maximum values for k_1, k_2, and k_3 were calculated from equations (8.8), (8.9) and (8.10) by substitution of appropriate input parameter values.

The ranges assumed for the principal independent parameters in evaluating the values for the rate constants are summarized in Table 9.1. The remaining independent parameters not listed in the table were assumed to be invariant and were set equal to the nominal values previously specified.

The distributions for the rate constants were assumed to be lognormal, typical of many environmental parameters. The minimum and maximum values given on Table 9.2 represent the 2.5 and 97.5 percentiles (i.e. approximately the range encompassed by ±2 standard deviations about the mean values). The mean values reported for the rate constants, which were calculated assuming lognormal distributions, agree closely with the nominal values given previously. This agreement suggests that the assumption that the distributions are lognormal is satisfactory.

As noted on Table 9.2, the rate constant for release from the sediment back to the water column (k_2) is calculated as a function of the parameter k_1. In a real system, k_2 will be correlated with k_1, hence it has been expressed as an explicit function of the root variable:

$$k_2 = \beta k_1$$

Values for the parameter β were calculated using extreme values calculated for k_1 and k_2 from equation (8.8) and (8.9). Like the rate constants, the parameter β was described as having a lognormal distribution.

The values determined above for the rate constants k_1 and k_3 and for the parameter β, were input to the mainstream probabilistic assessment. Results of the mainstream activities are discussed in Volume 1 of this report. In addition, the values reported above were used in a separate investigation of the use of a multi-component model for predicting the dynamics of Ra-226 behaviour in the reference lake. The results of the latter investigation are presented in a separate discussion paper included in this volume of the report.

REFERENCES

Gardner, R.H., Blaylock, B.G., and Hoffman, F.O., 1985. "A Dynamic Model of Radium-226 Behaviour in an Oligotrophic Lake." Prepared for SENES Consultants Limited on behalf of the National Uranium Tailings Program, CANMET, EMR, Ottawa, (in this report).

Kipphut, G.W., 1978. "An Investigation of Sedimenting Processes in Lakes." Ph.D. Thesis, Columbia University.

Kemp, A.L.W. and Harper, N.S., 1977. "Sedimentation Rates in Lake Huron and Georgian Bay." J. Great Lakes Research, Vol. 3 (3-4), pp. 215-220.

Baily, J.E., and Ollis, D.F., 1977. "Biochemical Engineering Fundamentals." McGraw Hill Book Company, Toronto.

Higbie, R. 1935. Transactions of AIChE, Vol. 35, pp. 365-372.

SENES Consultants Limited and Saskatchewan Research Council (SRC), 1984. "A Study of the Distribution of Naturally Occurring Radionuclides in Freshwater Benthos and Their Environment - Phase I." Research Report to the National Uranium Tailings Program, CANMET, EMR, Ottawa, Under Supply and Services Canada Contract Number OSQ83-00241.

Swanson, S.M., 1985. "A Study of the Distribution of Naturally Occurring Radionuclides in Freshwater Benthos and Their Environment - Phase II.: Prepared by SENES Consultants Limited and Saskatchewan Research Council for the National Uranium Tailings Program, CANMET, EMR, Ottawa. Under Supply and Services Canada Contract Number OSQ83-00241.

Wetzel, R.G., 1975. "Limnology." Saunders College Publishing; Philadelphia.

Table 9.1

RANGE OF INPUT PARAMETER VALUES FOR ASSESSING THE RATE CONSTANTS

Symbol	Description	Minimum	Mean	Maximum
σ_o	flux of sedimenting organic matter ($g.m^{-2}.yr^{-1}$)	50	100	200
σ_i	flux of sedimenting inorganic matter ($g.m^{-2}.yr^{-1}$)	25	80	160
σ_s	flux of total sedimenting matter ($g.m^{-2}.yr^{-1}$)	75	180	360
k_D^o	distribution coefficient for organic material ($m^3.g^{-1}$)	$2.4x10^{-3}$	$12x10^{-3}$	$30x10^{-3}$
k_D^i	distribution coefficient for inorganic material ($m^3.g^{-1}$)	$7.0x10^{-4}$	$3.5x10^{-3}$	$8.5x10^{-3}$
k_D^s	average distribution coefficient ($m^3.g^{-1}$)	$1.6x10^{-3}$	$8.2x10^{-3}$	$20x10^{-3}$
$\sigma_s k_D^s$	radionuclide transfer to sediment ($m.yr^{-1}$)	$1.2x10^{-1}$	1.47	7.2
σ_D	flux of stable (decayed) sediment ($g.m^{-2}.yr^{-1}$)	35	104	195
k_D^D	distribution coefficient for stabilized sediment ($m^3.g^{-1}$)	$1.1x10^{-3}$	$5.46x10^{-3}$	$13.4x10^{-3}$
$\sigma_D k_D^D$	radionuclide transfer to deep sediment ($m.yr^{-1}$)	$3.85x10^{-2}$	$5.68x10^{-1}$	2.6
k_L	mass transfer coefficient ($m.yr^{-1}$)	1.8	2.65	14.2

Table 9.2

RATE CONSTANT DISTRIBUTIONS FOR RADIUM REMOVAL IN THE LAKE SEDIMENT MODEL

Rate Constant	Distribution Type	Mean Value	Geo. Std. Dev.	Min. Value	Max. Value
k_1 rate of Ra-226 removal from the water column (yr^{-1})	lognormal	0.43	1.85	0.13	1.47
β * parameter relating k_2 to k_1	lognormal	0.18	1.35	0.10	0.33
k_3 rate of Ra-226 burial (loss) to deep sediment (yr^{-1})	lognormal	0.022	1.20	0.015	0.032

* $k_2 = \beta k_1$ = rate of Ra-226 release from sediment back to the water column (yr^{-1})

DISCUSSION PAPER ON

A DYNAMIC MODEL OF RADIUM-226 BEHAVIOUR IN AN OLIGOTROPHIC LAKE

by

R.H. Gardner, B.G. Blaylock, and F.O. Hoffman

for
SENES Consultants Limited
June, 1985

ABSTRACT

A multi-component model for predicting the dynamics of Ra-226 in a hypothetical oligotrophic lake situated in the Precambrian Shield region of Canada is described and time-dependent solutions are obtained for (1) a simulation of a clean lake with a contaminated input and (2) a simulation of a contaminated lake with a 99.9% cleanup of the input. Deterministic calculations, sensitivity and uncertainty analyses were performed for each case. The results show that: (1) time-dependent concentrations of Ra-226 in the flesh of game fish were most sensitive to changes in the rate constant describing the direct transfer of Ra-226 from water to game fish flesh; (2) equilibrium values of the water-to-game fish flesh concentration factor (CR) may give biased results when the aquatic system is not at equilibrium; and (3) a simpler model involving three components (water, sediment, game fish) can adequately predict the dynamic behaviour of radium in game fish. At equilibrium, Ra-226 concentrations predicted by the multi-component dynamic model are not appreciably different from those predicted using empirically derived water-to-game fish flesh CF's.

TABLE OF CONTENTS

Page

1.0 INTRODUCTION 1

2.0 METHODS 2
  2.1 Model Description 2
  2.2 Derivation of Rate Constants 2
  2.3 Description of Secondary Parameters 4
    2.3.1 Game Fish Flesh 5
    2.3.2 Forage Fish 7
    2.3.3 Vegetation 8
    2.3.4 Invertebrates 9
    2.3.5 Sediments 10
  2.4 Simulation Methods 11
  2.5 Sensitivity Analysis 11
  2.6 Uncertainty Analysis 12

3.0 RESULTS 13
  3.1 Scenario 1: Simulation to Equilibrium 13
  3.2 Scenario 2: 99.9% Cleanup 15

4.0 DISCUSSION AND RECOMMENDATIONS 17

REFERENCES 20

LIST OF TABLES

Follows Page

2.1 Assumed Physical and Biological Description of a Reference Oligotrophic Canadian Shield Lake 2

2.2 Assumed Deterministic Values for Concentration Factors and Turnover Times for Biota in a Reference Canadian Shield Lake 4

2.3 Nominal Values for Aquatic Model Parameters 4

2.4 Distributions for Aquatic Model Parameters 12

3.1 Sensitivity and Uncertainty Analysis - Concentration of Radium-226 in Game Fish Flesh: Scenario 1 14

3.2 Geometric Means and Geometric Standard Deviations of Model Predictions for Scenario 1 15

3.3 Sensitivity and Uncertainty Analysis - Concentration of Radium-226 in Game Fish Flesh: Scenario 2 16

3.4 Geometric Mean and Geometric Standard Deviations of Model Predictions for Scenario 2 17

LIST OF FIGURES

Follows Page

2.1 Illustration of a Multi-Component Model of an Oligotrophic Lake Showing 23 Pathways for Ra-226 Transfer 2

3.1 Changes in Water, Sediment and Game Fish Flesh Concentrations with Time, Starting with an Uncontaminated Lake at Time Zero 13

3.2 Change with Time of the Water-to-Game Fish Concentration Factor (CF), Starting with an Uncontaminated Lake at Time Zero 14

3.3 Changes in Water, Sediment and Game Fish Flesh Concentrations with Time, After the Amount of Radium in the Input Source has been Cleaned Up by 99.9% 15

3.4 Change with Time of the Water-to-Game Fish Concentration Factor (CF), After 99.9% of the Radium in the Input is Removed 16

1.0 INTRODUCTION

For the purposes of radiological assessment, the prediction of the concentration of a radionuclide in fish flesh is usually made using the mathematically simple concentration factor approach (Ophel 1978, National Council on Radiation Protection and Measurements 1984). The concentration factor (CF) assumes that concentrations in fish flesh and water are in equilibrium and that the concentration in fish is directly proportional to the concentration in water. All pathways and mechanisms of radionuclide uptake and elimination by and from fish, including food chain bioaccumulation and predation, are assumed to be implicitly included within empirical determinations of the CF. The concentration factor is obtained as a ratio of the amount of the radionuclide in a gram of fresh weight fish flesh (C_f) to the amount of the radionuclide in a mL of unfiltered water (C_w). Thus,

$$C_f = C_w \times CF$$

Criticism of the CF approach is that it is too simplistic, that ecological details of water, sediment, and food chain interaction should be addressed explicitly within the mathematical model, and that the time-dependent dynamics of the transfer of the radionuclide among the various components of an aquatic system cannot be explained by a factor derived under conditions of assumed equilibrium. Furthermore, there is the contention held by some that a more detailed model will enhance understanding of the aquatic system and therefore will produce results that are more accurate and defensible than the simple CF approach.

In recognition of these criticisms, the objectives of this paper are to construct a multi-component dynamic model for a hypothetical oligotrophic lake in the Precambrian Shield region of Canada, and to compare the predictions of the concentration of

Ra-226 in the flesh of game fish against results obtained using reported values of CF for Ra-226. It is assumed that this hypothetical lake is subject to contamination by Ra-226 through the disposal and/or leaching of uranium mill tailings from uranium mining and milling operations. Calculations are made for two scenarios: (1) where Ra-226 enters an uncontaminated lake and (2) where Ra-226 is removed from water entering the lake after Ra-226 concentrations in all lake components have reached equilibrium. For both scenarios, the multi-component model is used to calculate the concentrations of Ra-226 in the lake as a function of time. Uncertainties in model predictions are simulated by estimating uncertainties of model parameters (rate constants and the water-to-game fish CF) and translating these uncertainties into predicted concentrations of Ra-226 using both analytical and numerical error propagation techniques (Hoffman and Gardner 1983).

2.0 METHODS

2.1 Model Description

The multi-component model is composed of six compartments: water, sediment, game fish flesh, forage fish, invertebrates, and aquatic vegetation. There are twenty-three pathways of flow of radium among all model components (Figure 2.1). The lake is assumed to be a reference lake of the Canadian Precambrian Shield. The lake is a generalized oligotrophic lake that is cold, low in productivity and high in dissolved oxygen. The physical and biological characteristics of the lake are described in Table 2.1.

2.2 Derivation of Rate Constants

The most difficult task in constructing the multi-component model

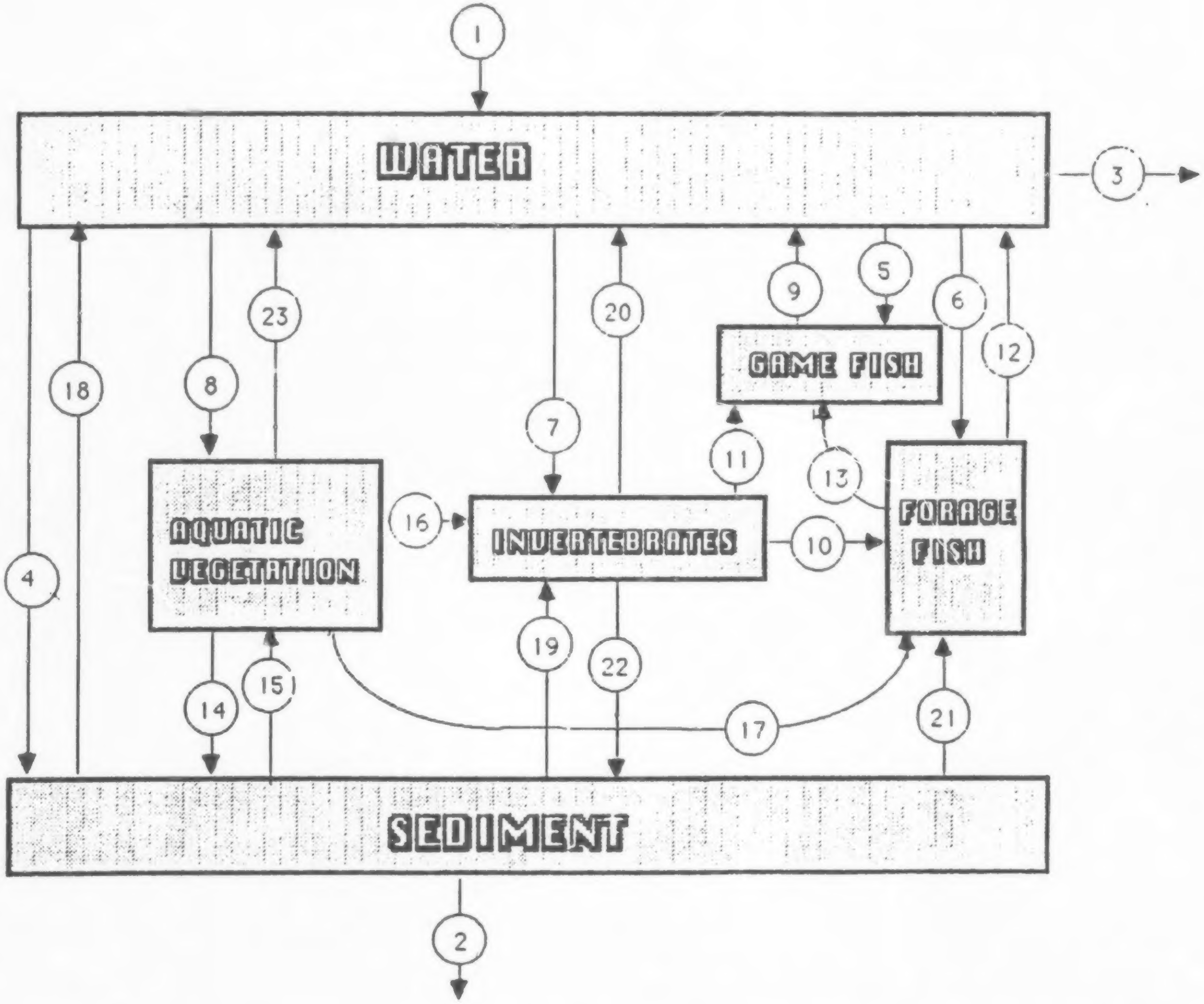
Table 2.1

ASSUMED PHYSICAL AND BIOLOGICAL DESCRIPTION OF A REFERENCE OLIGOTROPHIC CANADIAN SHIELD LAKE

Physical Parameters	Value
Surface area	160 ha
Average depth	10 m
Volume of water	1.6E10 L
Mass of the exchangeable sediment (assumed depth 5 cm)	1.4E7 kg
Biological Parameters	
Game fish flesh	500 kg
Forage fish whole	1600 kg
Aquatic vegetation	2.0E4 kg
Invertebrates	1.1E4 kg

FIGURE 2·1

ILLUSTRATION OF A MULTI-COMPONENT MODEL
OF A OLIGOTROPHIC LAKE SHOWING 23 PATHWAYS FOR ^{226}Ra TRANSFER



is quantitatively estimating the rate constants for the 23 flow pathways depicted in Figure 2.1. Data from experimental investigations are simply not available to directly quantify these parameters. Therefore, it was necessary to derive estimates of rate constants using collateral information and judgement. In the absence of specific determinations of rate constants, derived estimates depended on the following types of secondary parameters: (1) concentration factors for the equilibrium relationship between the concentration of Ra-226 in water and the concentration of Ra-226 in the individual model components, (2) the masses of the model components, (3) component turnover times (mean residence times of radium in the model components) and (4) the percentage, at equilibrium, of the total amount of radium per unit time entering or leaving a given model component that can be attributed to a specific flow pathway.

Information on concentration factors and component masses allowed estimates to be made of the concentrations and amounts of radium in each model component at equilibrium. Once the amount of radium in each model component was established, information on mean residence times for the model components permitted estimation of the equilibrium rates at which radium would enter and leave a model component (at equilibrium the total amount of radium lost per unit time must equal the rate of gain). Knowledge of the percentage contribution of a specific flow pathway to the total equilibrium rates of radium input to or losses from a model component permits an estimate of the equilibrium rate of radium flow (pCi/d) for that pathway.

After initial estimates of equilibrium flow rates were made for all 23 parameters, the amount of radium flowing through the entire system was brought into exact balance by adjusting the largest flow rates associated with each model component. This procedure was necessary to prevent errors from occurring due to mass imbalance during deterministic model simulations. After mass balance was achieved, rate constants (d^{-1}) were derived by dividing each equilibrium flow rate by the amount of radium in the model component from which the flow originated.

The concentration factors and mean residence times assumed for estimating the deterministic values of the rate constants are presented in Table 2.2. Deterministic estimates of the 23 rate constants and their associated equilibrium flow rates are given in Table 2.3. The information sources and assumptions for estimating the secondary parameters used to derive the rate constants are described in the following section.

2.3 Description of Secondary Parameters

When data were available values for the secondary parameters were based on the mean or geometric mean of data representative of an oligotrophic lake. When specific data were not available, the best estimates of the secondary parameters were derived from general information or from data for an analogue radionuclide. Some of the parameters were difficult to estimate because the model components consisted of several groups of organisms. For example, the invertebrate component consisted of the zooplankton and benthic invertebrates; thus the biomass and turnover time for the compartment should reflect an average for both groups of organisms. In addition, the biomass of the different food chain components and their associated turnover times were for daily annual averages.

Although a considerable amount of information is available for the summer months for some organisms, such as phytoplankton for the summer months, there is a paucity of data for biomass and turnover times during the winter months when the lake is covered with ice. As a result, many of the parameters are based on

Table 2.2

ASSUMED DETERMINISTIC VALUES FOR CONCENTRATION FACTORS AND TURNOVER TIMES FOR BIOTA IN A REFERENCE CANADIAN SHIELD LAKE

Model Component	Concentration Factor (kg/L)	Turnover Time (d or Y)
Game fish flesh	16	8 y
Forage fish whole	180	365 d
Aquatic Vegetation	1500	5 d
Invertebrates	1200	38 d
Water	1	365 d
Sediment	4000*	8.2 y

* K_d value for exchangeable sediment

Table 2.3

NOMINAL VALUES FOR AQUATIC MODEL PARAMETERS

Rate Constant		Equilibrium Rates	Rate Constants**
Number	Name*	(pCi per day)	(per day)
1	Input	4.72 E 7	4.72 E 7
2	Deeploss	3.36 E 6	6.00 E-5
3	Outflow	4.38 E 7	2.74 E-3
4	Water-to-Sed	1.87 E 7	1.17 E-3
5	Water-to-Game	2.44 E 0	1.52 E-10
6	Water-to-Forg	7.10 E 2	4.44 E-8
7	Water-to-Inv	3.14 E 5	1.96 E-5
8	Water-to-Av	5.90 E 6	3.69 E-4
9	Game-to-Water	2.72 E 0	3.40 E-4
10	Inv-to-Forg	3.95 E 1	2.74 E-6
11	Inv-to-Game	1.10 E-1	7.64 E-9
12	Forg-to-Water	7.89 E 2	2.74 E-3
13	Forg-to-Game	1.64 E-1	5.70 E-7
14	Av-to-Sed	3.00 E 1	1.00 E-6
15	Sed-to-Av	3.00 E 4	5.35 E-7
16	Av-to-Inv	4.90 E 4	1.63 E-3
17	Av-to-Forg	1.96 E 1	6.53 E-7
18	Sed-to-Water	1.53 E 7	2.74 E-4
19	Sed-to-Inv	7.49 E 3	1.34 E-7
20	Inv-to-Water	3.70 E 5	2.57 E-2
21	Sed-to-Forg	1.96 E 1	3.50 E-10
22	Inv-to-Sed	1.44 E 2	1.00 E-5
23	Av-to-Water	5.88 E 6	1.96 E-1

* The abbreviations for parameter names:

Sed - Sediment	Forg - Forage Fish
Inv - Invertebrates	Game - Game Fish
Av - Aquatic Vegetation	

** The rate constants were derived by dividing each equilibrium flow rate by the amount of radium in the model component from which the flow originated, for example:

$$\frac{[\text{Water-to-game equilibrium rate of 2.44E0 pCi.day}^{-1}]}{[\text{Water radium mass of 1.6E10 L x 1.0 pCi.L}^{-1}]}$$

$$= [\text{Water-to-game rate constant of 1.52E-10 day}^{-1}]$$

Table 2.3

NOMINAL VALUES FOR AQUATIC MODEL PARAMETERS

Rate Constant		Equilibrium Rates (pCi per day)	Rate Constants** (per day)
Number	Name*		
1	Input	4.72 E 7	4.72 E 7
2	Deeploss	3.36 E 6	6.00 E-5
3	Outflow	4.38 E 7	2.74 E-3
4	Water-to-Sed	1.87 E 7	1.17 E-3
5	Water-to-Game	2.44 E 0	1.52 E-10
6	Water-to-Forg	7.10 E 2	4.44 E-8
7	Water-to-Inv	3.14 E 5	1.96 E-5
8	Water-to-Av	5.90 E 6	3.69 E-4
9	Game-to-Water	2.72 E 0	3.40 E-4
10	Inv-to-Forg	3.95 E 1	2.74 E-6
11	Inv-to-Game	1.10 E-1	7.64 E-9
12	Forg-to-Water	7.89 E 2	2.74 E-3
13	Forg-to-Game	1.64 E-1	5.70 E-7
14	Av-to-Sed	3.00 E 1	1.00 E-6
15	Sed-to-Av	3.00 E 4	5.35 E-7
16	Av-to-Inv	4.90 E 4	1.63 E-3
17	Av-to-Forg	1.96 E 1	6.53 E-7
18	Sed-to-Water	1.53 E 7	2.74 E-4
19	Sed-to-Inv	7.49 E 3	1.34 E-7
20	Inv-to-Water	3.70 E 5	2.57 E-2
21	Sed-to-Forg	1.96 E 1	3.50 E-10
22	Inv-to-Sed	1.44 E 2	1.00 E-5
23	Av-to-Water	5.88 E 6	1.96 E-1

* The abbreviations for parameter names:

Sed - Sediment
Inv - Invertebrates
Av - Aquatic Vegetation
Forg - Forage Fish
Game - Game Fish

** The rate constants were derived by dividing each equilibrium flow rate by the amount of radium in the model component from which the flow originated, for example:

$$\frac{[\text{Water-to-game equilibrium rate of 2.44E0 pCi.day}^{-1}]}{[\text{Water radium mass of 1.6E10 L x 1.0 pCi.L}^{-1}]}$$

$$= [\text{Water-to-game rate constant of 1.52E-10 day}^{-1}]$$

general knowledge and best judgement. Parameters for the physical characteristics of the reference oligotrophic lake, surface area and mean depth (Table 2.1), were specified by SENES Consultants Limited (Halbert 1985).

Development of parameter values for the biological components of the model (game fish flesh, forage fish, invertebrates, and aquatic vegetation) and sources of information are discussed in the following sections.

2.3.1 Game Fish Flesh

The estimated range of productivity for game fish in an oligotrophic lake is 0.26 to 2.5 kg per ha per year (Swanson 1985a). For the reference lake a productivity of 1 kg per ha per year was assumed. The productivity multiplied by the average turnover time (or replacement time) and lake area will give an estimate of the biomass (sometimes referred to as standing crop) of the model component. Swanson (1983) reported a wide age range of 5 to 36 years for lake trout collected from two oligotrophic lakes. Swanson (1985a) estimated that the average age of lake trout in an oligotrophic lakes was between 7 to 16 years depending upon whether the lake has supported commercial fishing. For the reference lake an average age of eight years for game fish was assumed. The biomass of game fish in the 160 ha lake is 1280 kg (=1 kg/(ha*yr)*160 ha*8yr).

The compartment of interest in our model is game fish flesh or the edible portion of the fish. Rope and Whicker (1980) reported that 38% of the wet weight of rainbow trout was flesh. Using this percentage as a conversion factor the biomass of the game fish flesh component is 486 kg or rounded to 500 kg (Table 2.1).

The concentration factor (CF) for the game fish flesh component of the model was assumed to be 16. A review of recent data (Halbert et al. 1985) indicated that the concentration factors for game fish flesh ranged from 1 to 603, and that these values were lognormally distributed with a geometric mean of 15.7. A study by Rope and Wicker (1985) that was not included in the above review reported concentration factors of 0.3 to 3.5 for rainbow trout; however, these values are at the lower end of the range of concentration factors reported previously.

Radium is a bone seeker and moves rapidly to the bone from the other tissues of fish. Rope and Whicker (1980) reported a maximum bone/flesh ratio of 250 for rainbow trout. Although uptake and elimination studies have not been conducted for Ra-226 in fish, it is assumed that the residence time for radium in fish is similar to Sr and Ca (Nelson 1966), and that the turnover rate of the fish population is much faster than the biological half-life of Ra-226 in fish. The estimated average age of the game fish population which is 8 yr would be the turnover or residence time of Ra-226 in the game fish component of the model.

Game fish can obtain Ra-226 directly from the water or through the food chain. Because the concentration factor for the game flesh and forage fish flesh are essentially the same (Halbert et al. 1985, Swanson 1983) it was assumed that most (90%) of the body burden of Ra-226 came from the water component of the reference aquatic ecosystem. It was assumed that the remaining 10% of the body burden was taken in via the food chain, 6% from the forage fish component and 4% from the invertebrate component. Although the diet of game fish consists primarily of forage fish, the invertebrates have a much higher concentration factor than the forage fish; thus they have a much higher concentration per unit weight.

2.3.2 Forage Fish

The forage fish component is made up of dace, shiners, chubs and small white fish (Swanson 1983, Hesslein and Slavicek 1984). For this model component the biomass consists of the whole fish. We assume the entire fish is consumed by the predacious game fish. The estimate of the biomass for the forage fish component was based on the relationship between the biomass of forage fish to the biomass of the other compartments in the oligotrophic lake, and the assumption that the average turnover time for forage fish was one year. Based on these assumptions the estimated biomass of the forage fish compartment is 1600 kg.

The concentration factors for forage fish flesh ranged from 3 to 90 with a geometric mean of 12.8 (Halbert et al. 1985); however, for the forage fish component a concentration factor is required for the whole fish. The concentration factors reported for whole forage fish by Swanson (1983) and Hesslein and Slavicek (1984) ranged from 53 to 1473. Based on these data a concentration factor of 180 was assumed for forage fish. The turnover time of the forage fish population which was assumed to be one year, would determine the turnover of Ra-226 in this component of the model because of the long biological half-life of Ra-226 in fish.

Since game fish acquire 90% of their body burden of Ra-226 directly from water, it is logical to assume that approximately the same percentage would be acquired by forage fish. The remaining 10 percent of the forage fish body burden would be taken up via the food chain. Based on stomach content analysis for forage fish reported by Swanson (1985b), it was assumed that 5% of the Ra-226 body burden came from the invertebrate component, 2.5% came from the vegetation component, and the remaining 2.5% from the sediment component.

2.3.3 Vegetation

The vegetation component is composed of phytoplankton, periphyton and macrophytes. The reference lake is assumed to be low in productivity with phytoplankton being the prime food base for the higher organisms. There is a large volume of literature on phytoplankton productivity and biomass. Unfortunately much of it does not apply to the Canadian Shield lakes, and the situation is further complicated by the fact that biomass is measured as chlorophyll, carbon, dry weight or wet weight. In most cases, conversions were necessary to obtain wet weight measurements. The biomass of phytoplankton for oligotrophic lakes was estimated to range from 1.0 to 20 mg/L wet wt (Kimmel and Soballe 1985). Schindler and Holmgren (1971) reported that the biomass of phytoplankton, dominated by Chrysophyceae, ranged from 0.12 to 5.4 mg/L and from 2.1 to 13.4 mg/L for July and August for four lakes in the Experimental Lake Area. The turnover time reported by Schindler and Holmgren (1971) ranged from 0.44 to 2.77 d. Because the turnover time for phytoplankton is reduced during the winter, the turnover time of 5 days was assumed for the reference lake. These biomass measurements and the turnover times listed above are in agreement with the data for several oligotrophic lakes listed by Wetzel (1975). The estimated biomass for phytoplankton is 1.25 mg/L wet wt for the reference lake; thus the biomass for the vegetation component is 2.0E4 kg.

The vegetation component of the lake ecosystem is dominated by phytoplankton. Rope and Whicker (1980) reported concentration factors for algae ranging from 1100 to 17,500 and 960 to 6900 for plants. Hesslein and Slavicek (1984) determined concentration factors for Ra-226 in macrophytes that ranged from 110 to 5000. The range of concentration factors for radium reported for phytoplankton, algae, and macrophytes in the Animas River in Colorado-New Mexico was 400-1200, 200-2800 and 725-1200, respectively (Anderson et al. 1963). The concentration factors selected for the vegetation component of the model was 1500. It was assumed that the turnover rate of Ra-226 in the vegetation component was determined by the relatively short turnover time of the populations in the vegetation components.

2.3.4 Invertebrates

The invertebrates component of the reference lake is composed of zooplankton and benthic invertebrates. The benthic invertebrates are dominated by chironomids and other aquatic insects, oligochets, and clams. The density of the invertebrates ranged from 1.7 to 4.7 g wet wt/m^2 (Swanson 1985a) and are dominated by zooplankton. The dominant zooplankters are rotifers, copepods, and diaptomus. Estimates of the density of zooplankton in a oligotrophic lake were taken from Wetzel (1975) and the Ontario Water Resources Commission (1972). The average biomass for the invertebrate compartment was estimated as 1.1 E 4 kg. According to Wetzel estimated turnover time for zooplankton in oligotrophic lakes ranges from 2 to 80 days. The turnover time for the invertebrate compartment in the reference lake was estimated as 38 days.

In a review by Blaylock (1982) radium concentration factors for invertebrates ranged from 500 to 1000. A research report by Swanson (1985b) contains the most comprehensive data on concentration factors for radium for benthic invertebrates that is currently available. These data, reporting concentration factors for a variety of invertebrates were reviewed by Halbert (1985). Concentration factors for these data ranged from 45 to 43,333 with a geometric mean concentration of 1385.

Since the invertebrate component is dominated by zooplankton, and since we are not aware of any information for the concentration factors for zooplankton, a concentration factor of 1200 was selected to represent both zooplankton and benthic invertebrates. This is slightly lower than the geometric mean concentration factor reported by Halbert (1985); however, the benthic invertebrates are closely associated with sediments and their body burden of Ra-226 also includes the G.I. tract content. In contrast most of the zooplankton is not associated with bottom sediment but accumulate their body burden directly from water. It was assumed that the turnover rate of Ra-226 of the invertebrate component was determined by the turnover time of the invertebrate populations.

2.3.5 Sediments

The exchangeable sediment component of the oligotrophic lake was assumed to be the top 5 cm of the sediment. This exchangeable sediment is composed of a high percentage of organic material and interstitial water in addition to organic material. The source of Ra-226 to the sediment is water, dissolved solids and organic material from the biotic populations. Radium-226 entering the exchangeable sediment can be transferred back to the water column or move deeper into the sediment (deeploss) where it is not readily available for remobilization and is considered lost to the biological community of the lake. Remobilization of Ra-226 occurs through chemical changes (i.e. changes in chemical form) or through bioturbation (i.e. biological activity of organisms). The values for the rate constants for the loss of Ra-226 from water to sediment, exchange of sediment to water, and the loss to deep sediment were provided by SENES Consultants Limited (Scharer and Halbert 1985).

2.4 Simulation Methods

The dynamics of the concentrations of Ra-226 in the model components illustrated in Figure 2.1 were mathematically represented by a coupled set of ordinary differential equations. The derivatives for each component were constructed as functions of the amount of Ra-226 in the particular component and the rate constants which describe the movement of material between components. Thus, the derivative describing the rate of change of Ra-226 in game fish flesh is:

$$\begin{aligned} dx/dt = {} & (\text{WATER*Water-to-Game}) + (\text{FORAGE FISH*Forg-to-Game}) + \\ & (\text{INVERTEBRATES*Inv-to-Game}) + (\text{GAME*Game-to-Wa}) \end{aligned}$$

where the components illustrated in Figure 2.1 are in capital letters and the acronyms for the rate constants are the same as those listed in Table 2.3. The remaining equations for the derivatives can be easily derived from Figure 2.1 and Table 2.3.

The time dependent solutions to these equations is obtained by means of a fourth/fifth order Runge-Kutta, RKF45 (see Forsythe et al. 1977, for documentation). This method of representing and simulating an aquatic model was chosen because the dynamics of the system are stable and an adequate representation of time-averaged rate constants is provided. In addition, an aquatic model based on this formalism is possible to parameterize from available information.

2.5 Sensitivity Analysis

Sensitivity analyses of these equations were performed by perturbing each rate constant by 1% of their nominal value (Table 2.3), solving the system over the time interval of interest, and calculating the percent change in each component as a result of the 1% perturbation of a particular rate constant. The process was repeated by resetting the rate constant to its nominal value, perturbing the next rate constant and solving the system again. A 1% perturbation has been shown to provide an adequate numerical estimation of the analytical sensitivities of the equation (Gardner 1984).

2.6 Uncertainty Analysis

Estimates of the mean, variance and frequency distribution of model predictions as a function of the uncertainties associated with the model rate constants were obtained by Monte Carlo methods (Rubinstein 1981). Monte Carlo methods involve the iterative selection of random rate constants from their respective distributions, solving the system over the time period of interest (for this case, years 1,5, and 10), and saving the sets of random rate constants and model solutions. This process was iterated 100 times for each model scenario and the results statistically analyzed. In order to accomplish this process, frequency distributions must be specified for each model rate constant associated with uncertainty. As we have mentioned previously in other papers (SENES 1984, Hoffman and Gardner 1983, Hoffman et al. 1985) these distributions do not represent the expected variability of the parameters. Instead, these distributions are intended to encompass the extent of uncertainty associated with parameter estimation. Our estimates of these input distributions are listed in Table 2.4.

In Table 2.4, the shape and size of the selected distributions are primarily based on judgemental estimates. All sources of uncertainty are encompassed in these distribution estimates with the exception of uncertainties in the specification of biomass. Where sufficient data was available from which inferences could be drawn about the plausible shape and size of a distribution, a

Table 2.4

DISTRIBUTIONS FOR AQUATIC MODEL PARAMETERS

Rate Constant		Distribution	Mean or	Standard Dev.
Number	Name**	Type **	Minimum	or Maximum
1	Input	Constant	4.72 E7	-
2	Deeploss	Lognormal	6.00 E-5	1.29 E 0
3	Outflow	Normal	2.74 E-3	2.74 E-4
4	Water-to-Sed	Lognormal	1.17 E-3	1.87 E 0
5	Water-to-Game	Lognormal	1.52 E-10	3.37 E 0
6	Water-to-Forg	Lognormal	4.44 E-8	3.37 E 0
7	Water-to-Inv	Loguniform	6.13 E-7	6.27 E-4
8	Water-to-Av	Loguniform	6.14 E-5	2.21 E-3
9	Game-to-Water	Uniform	1.70 E-4	6.80 E-4
10	Inv-to-Forg	Loguniform	8.67 E-7	8.66 E-6
11	Inv-to-Game	Loguniform	7.64 E-10	7.64 E-8
12	Forg-to-Water	Loguniform	2.74 E-4	2.74 E-2
13	Forg-to-Game	Uniform	1.90 E-7	1.71 E-6
14	Av-to-Sed	Loguniform	1.00 E-7	1.00 E-5
15	Sed-to-Av	Loguniform	5.36 E-8	5.36 E-6
16	Av-to-Inv	Loguniform	5.17 E-4	5.16 E-3
17	Av-to-Forg	Loguniform	2.07 E-7	2.06 E-6
18	Sed-to-Water	Loguniform	2.74 E-5	2.74 E-3
19	Sed-to-Inv	Loguniform	1.33 E-8	1.33 E-6
20	Inv-to-Water	Loguniform	2.57 E-3	2.57 E-1
21	Sed-to-Forg	Loguniform	3.50 E-11	3.50 E-9
22	Inv-to-Sed	Loguniform	1.00 E-6	1.00 E-4
23	Av-to-Water	Loguniform	1.96 E-2	1.96 E 0

* The abbreviations for parameter names:

Sed - Sediment
Inv - Invertebrates
Av - Aquatic Vegetation
Forg - Forage Fish
Game - Game Fish

** The means and standard deviations are listed for the normal distributions, for the lognormal distribution the values are geometric means and geometric standard deviations (the geometric standard deviation is unitless). The minimum and maximum values are given for the uniform and loguniform distributions.

normal or lognormal distribution was specified. Where large uncertainties existed in our ability to quantify a rate constant, and where data were insufficient to suggest a distribution shape and size, a loguniform distribution was selected. In those cases with limited data where only moderate uncertainties were expected, uniform distributions were selected.

Loguniform distributions ranged from one to three orders of magnitude, with the majority set at two orders of magnitude. The ranges of uniform distributions were less than one order of magnitude. The values for the distributions assigned to the rate constants: Deeploss, Water-to-Sed, and Sed-to-Water, came from estimates provided directly by SENES Consultants Limited (Scharer and Halbert 1985). The deterministic estimates for these rate constants are equivalent to the arithmetic means of the normal and uniform distributions and the geometric means of the lognormal and loguniform distributions.

3.0 RESULTS

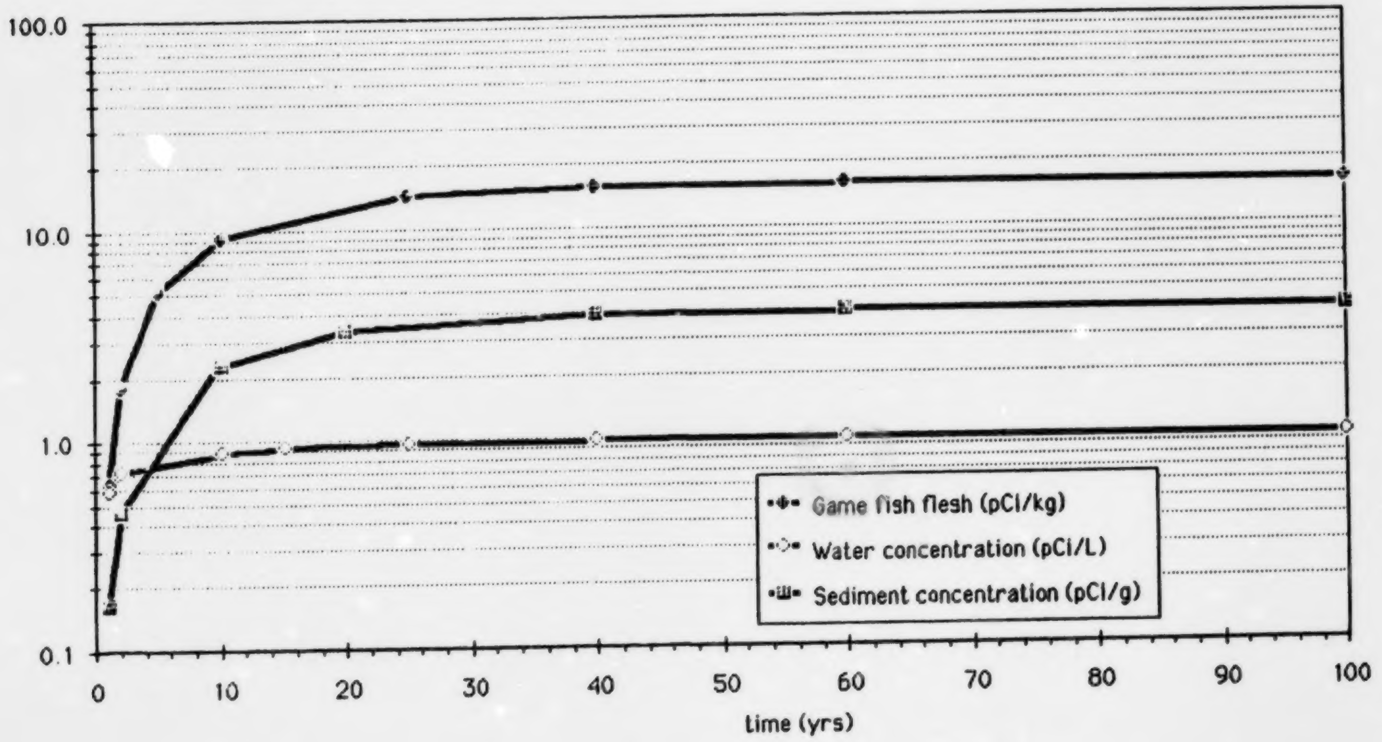
Two scenarios were simulated with the aquatic model: (1) Ra-226 was put into an uncontaminated aquatic system until equilibrium was reached in all model components, and (2) the effect of a cleanup of a contaminated source (a 99.9% reduction of the daily input rate of radium) in a system that is already at the equilibrium values established in Scenario 1.

3.1 Scenario 1: Simulation to Equilibrium

The deterministic results (i.e. all rate constants set at their nominal values, Table 2.3) of the first scenario are illustrated in Figure 3.1. The dynamics of the system are a direct function of the mean residence times of radium in each model component (the mean residence time is the reciprocal of the sum of the rate

FIGURE 3·1

CHANGES IN WATER, SEDIMENT AND GAME FISH-FLESH CONCENTRATIONS WITH TIME, STARTING WITH AN UNCONTAMINATED LAKE AT TIME ZERO



constants of the outputs for a given model component). Water, with a mean residence time of 0.63y, reached equilibrium relatively quickly. The amount of radium in game fish flesh (mean residence time = 8y) takes considerably longer to approach equilibrium. The active layer of sediment has a mean residence time similar to game fish (8.2 y) and approaches equilibrium in a similar fashion.

Figure 3.2 illustrates the dynamic changes of the concentration ratio of water-to-game fish flesh (CF). The equilibrium CF is equal to 16.0. At year 8 of the simulation (Figure 3.2) CF=8.9, at year 16 CF=12.8 and at year 24 CF=14.5. Thus it takes about 3 times the mean residence time of game fish for the concentration factor to be within 90% of the equilibrium value. All of these values, however, are well within the reported range of measured CF's.

The sensitivities of the concentration factor to the individual rate constants change as a function of time (Table 3.1). Initially, the concentration of Ra-226 in game fish flesh is most affected by the rate constant water-to-game. For example, a 1% change in the rate constant results in a 0.95% increase in the fish flesh concentration. The initial concentration of Ra-226 in game fish flesh is also somewhat affected by the loss of water (outflow) from the lake. At years 5 and 10 of the simulation, the sensitivity of the Ra-226 concentration in game fish flesh to the water-to-game transfer rate diminishes and other parameters become more important. In year 10 the game-to-water rate constant dominates.

The results of an uncertainty analysis can differ from those of a sensitivity analysis depending on the interactions among highly connected model components (Fig. 3.1) and differences in the uncertainties associated with the rate constants (Table 2.4).

Table 3.1

SENSITIVITY AND UNCERTAINTY ANALYSIS*: CONCENTRATION OF RADIUM-226 IN GAME FISH FLESH: SCENARIO 1

Rate Constant		Scenario 1: Initial Clean Case					
		Year 1		Year 5		Year 10	
Number	Name**	S	U	S	U	S	U
1	Input						
2	Deeploss					-0.28	
3	Outflow	0.12		0.11		-0.22	
4	Water-to-Sed					-0.17	
5	Water-to-Game	0.95	0.76	0.91	0.61	0.62	0.58
6	Water-to-Forg					-0.24	
7	Water-to-Inv					-0.25	
8	Water-to-Av					-0.28	
9	Game-to-Water			0.24		-0.74	
10	Inv-to-Forg					-0.29	
11	Inv-to-Game					-0.25	
12	Forg-to-Water						0.04
13	Forg-to-Game					-0.23	0.04
14	Av-to-Sed		0.05		0.04	-0.29	
15	Sed-to-Av					-0.29	
16	Av-to-Inv					-0.28	
17	Av-to-Forg					-0.29	
18	Sed-to-Water		0.04			-0.32	
19	Sed-to-Inv					-0.29	
20	Inv-to-Water					-0.33	
21	Sed-to-Forg					-0.29	
22	Inv-to-Sed						
23	Av-to-Water					-0.29	

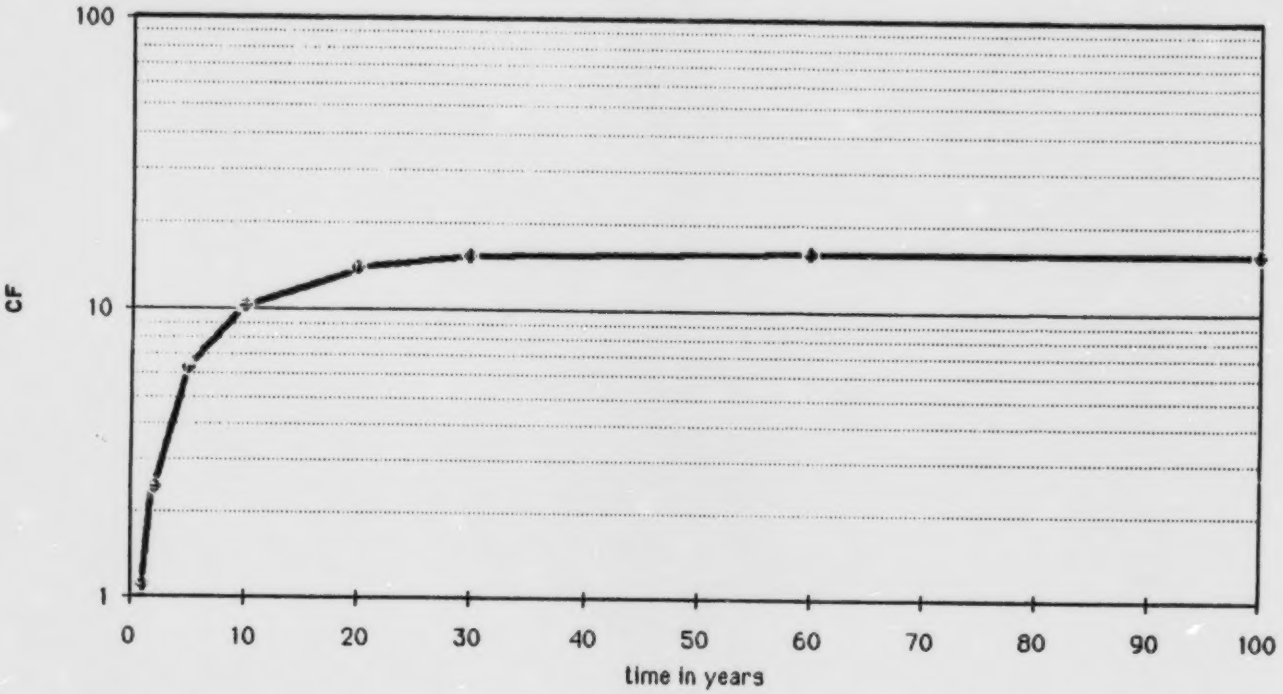
* Sensitivities, S, are calculated as the percent change in the model response due to a 1% decrease in a particular rate constant. The uncertainty index, U, is the square of the simple correlation coefficient of the log transformed model response with the log transformed values of the rate constant of interest calculated from 100 Monte Carlo simulations with all parameters varied as shown in Table 2.4. Values of S<0.1 and U<0.04 are not shown.

** The abbreviations for parameter names:

Sed - Sediment	Forg - Forage Fish
Inv - Invertebrates	Game - Game Fish
Av - Aquatic Vegetation	

FIGURE 3·2

CHANGE WITH TIME
OF THE WATER-TO-GAME FISH CONCENTRATION FACTOR (CF),
STARTING WITH AN UNCONTAMINATED LAKE AT TIME ZERO



These effects are apparent in Table 3.1 where several rate constants that were not identified as sensitive (aquatic vegetation-to-sediment and sediment-to-water, in year 1) have an effect on the uncertainty of the concentration of Ra-226 in game fish flesh. However, the most important rate constant identified by the sensitivity analysis (water-to-game), is also the most important rate constant in the uncertainty analysis. The water-to-game rate constant contributes 76%, 61% and 58% of the variability of the concentration of radium in game fish in year 1,5 and 10 respectively (Table 3.1).

The geometric means and geometric standard deviations of all model predictions for the equilibrium scenario are shown in Table 3.2. Among model components, water has the smallest uncertainty (geometric standard deviation ranges from 1.12 to 1.17) and invertebrates have the largest (7.69 to 8.62). The uncertainty of the water-to-game fish flesh concentration factor increases as a function of time (geometric standard deviation ranging from 3.10 to 3.30). As noted earlier, these results reflect only the uncertainties associated with model rate constants. Additional uncertainties in biomass estimates for model components (e.g. game fish flesh) have not been included. For this reason the estimated uncertainty in the water-to-game fish flesh concentration factor at equilibrium is less than that of actual measured values.

3.2 Scenario 2: 99.9% Cleanup

Scenario 2 involves a 99.9% reduction of the input of radium into a system whose components are at the equilibrium concentrations established by Scenario 1. The changes of radium concentration with time are illustrated in Figure 3.3. The entire system takes approximately 80 years to approach the new equilibrium values in spite of large differences in the mean residence times of each

Table 3.2

GEOMETRIC MEANS AND GEOMETRIC STANDARD DEVIATIONS OF MODEL PREDICTIONS FOR SCENARIO 1

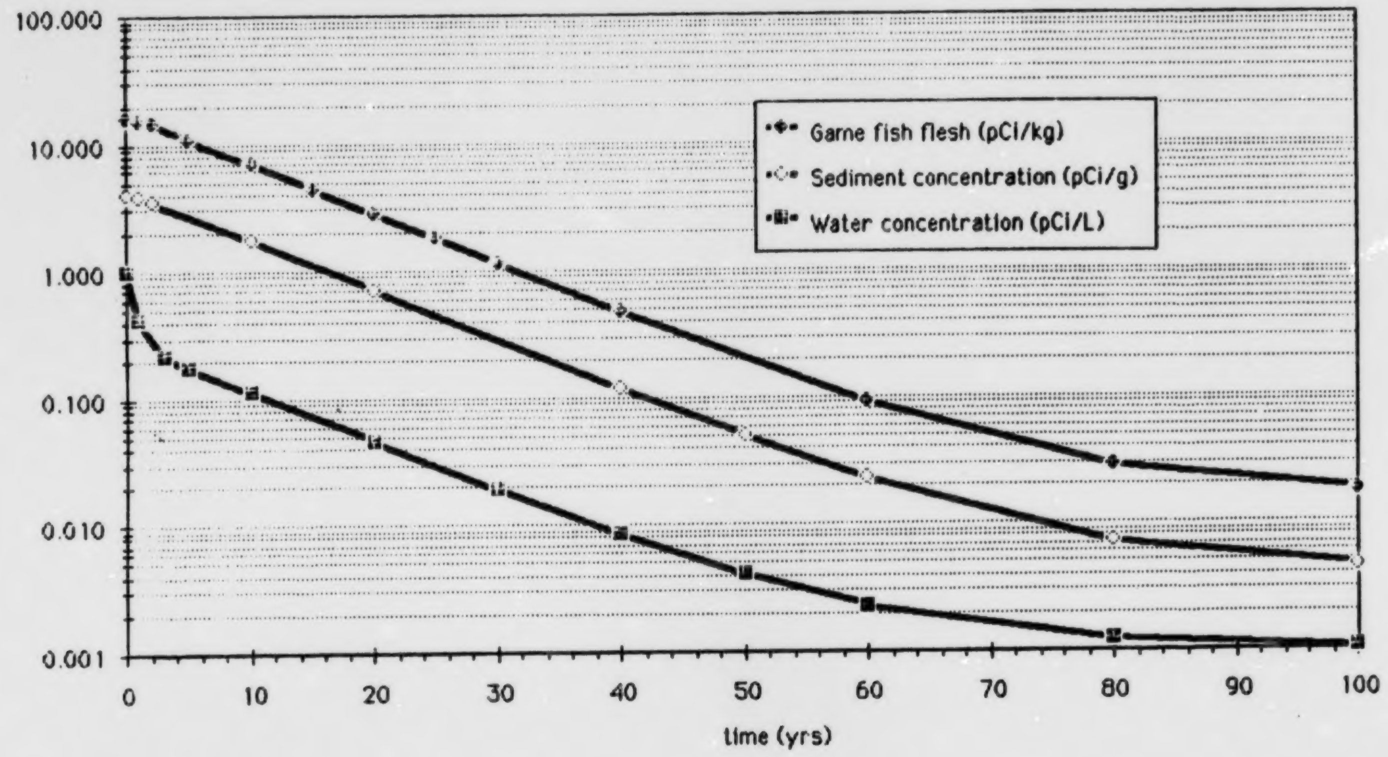
	Scenario 1: Initial Clean Case					
	Year 1		Year 5		Year 10	
Compartment	Mean	S.D.	Mean	S.D.	Mean	S.D.
Water*	8.95 E 9	1.12	1.28 E 10	1.17	1.37 E 10	1.15
Game Fish*	4.00 E 2	3.22	3.56 E 3	3.29	6.38 E 3	3.33
Forage Fish	6.40 E 4	3.05	2.20 E 5	4.38	2.74 E 5	4.91
Aquatic Vegetation*	1.85 E 7	4.47	2.72 E 7	4.53	2.93 E 7	4.51
Invertebrates*	9.33 E 6	7.69	1.68 E 7	8.62	1.84 E 7	8.48
Sediments*	2.18 E 9	1.80	1.68 E 10	1.64	3.02 E 10	1.65
CF: Water to Game Fish**	1.43 E 0	3.10	8.88 E 0	3.26	1.49 E 1	3.30

* The geometric mean values equal the mass of radium-226 in each model component in pCi.

** The concentration factor (CF) equals:
[pCi Ra-226/kg fresh fish flesh] ÷ [pCi Ra-226/L water]

FIGURE 3·3

CHANGES IN WATER, SEDIMENT
AND GAME FISH - FLESH CONCENTRATIONS WITH TIME,
AFTER 99·9% OF THE RADIUM IN THE INPUT IS REMOVED



component. The reason is that the aquatic system is affected by the large reservoir of radium in the sediment. For example, at year 5 of the simulation the transfer of radium from sediment-to-water equals 1.04E7 pCi per day (the rate constant, 2.75E-4 d^{-1}, multiplied by the amount of radium in the sediments at year 5, 3.84E10 pCi). This rate of transfer is a factor of 220 times greater than the daily input of radium (4.72E4 pCi per day) from outside the system. Thus the decline of radium in water is slowed and the loss of radium from game fish flesh is longer than anticipated by the mean residence time of radium in fish flesh.

The dynamics of the model components dramatically affect the water-to-game fish flesh concentration factor (Figure 3.4). The CF is equal to 16.0 at year 0 (the instant that the input rate of radium was reduced by 99.9%) and then rises rapidly to a high of 60.4 at year 10. The CF approaches 20.0 by year 80 and is near its initial value by year 100.

The most sensitive rate constants of Scenario 2 (Table 3.3) reflect the losses of radium from water (outflow); game fish flesh (game-to-water); and exchangeable sediments (sediment-to-water). The uptake of radium by game fish flesh is only moderately important in year 10, while the loss of radium from game fish to water becomes increasingly important with time, reaching a high of -1.06 in year 10 (i.e. a 1% increase in this rate constant results in a change of radium in game fish flesh greater than 1%).

The uncertainty analysis also shows the strong time dependence of this scenario (Table 3.3), but, as in Scenario 1, the important rate constants are not always the same as those of the sensitivity analysis. In year 1 the sediment-to-water rate constant accounts for 48% of the variability of the concentration of radium in game fish flesh. However, because of the large

Table 3.3

SENSITIVITY AND UNCERTAINTY ANALYSIS*: CONCENTRATION OF RADIUM-226 IN GAME FISH FLESH: SCENARIO 2

Rate Constant		Scenario 2: 99.9% Cleanup Case					
		Year 1		Year 5		Year 10	
Number	Name**	S	U	S	U	S	U
1	Input						
2	Deeploss						
3	Outflow	0.71		0.74		0.17	0.04
4	Water-to-Sed	0.28		0.14	0.10	0.73	
5	Water-to-Game		0.16	0.18	0.26	0.29	0.30
6	Water-to-Forg						
7	Water-to-Inv						
8	Water-to-Av						
9	Game-to-Water	-0.12		-0.58	0.08	-1.06	0.23
10	Inv-to-Forg						
11	Inv-to-Game						
12	Forg-to-Water						0.04
13	Forg-to-Game						
14	Av-to-Sed		0.06		0.08		0.07
15	Sed-to-Av						
16	Av-to-Inv						
17	Av-to-Forg						
18	Sed-to-Water	-0.40	0.48	-0.60	0.25	-0.20	
19	Sed-to-Inv						
20	Inv-to-Water						
21	Sed-to-Forg						
22	Inv-to-Sed						
23	Av-to-Water						

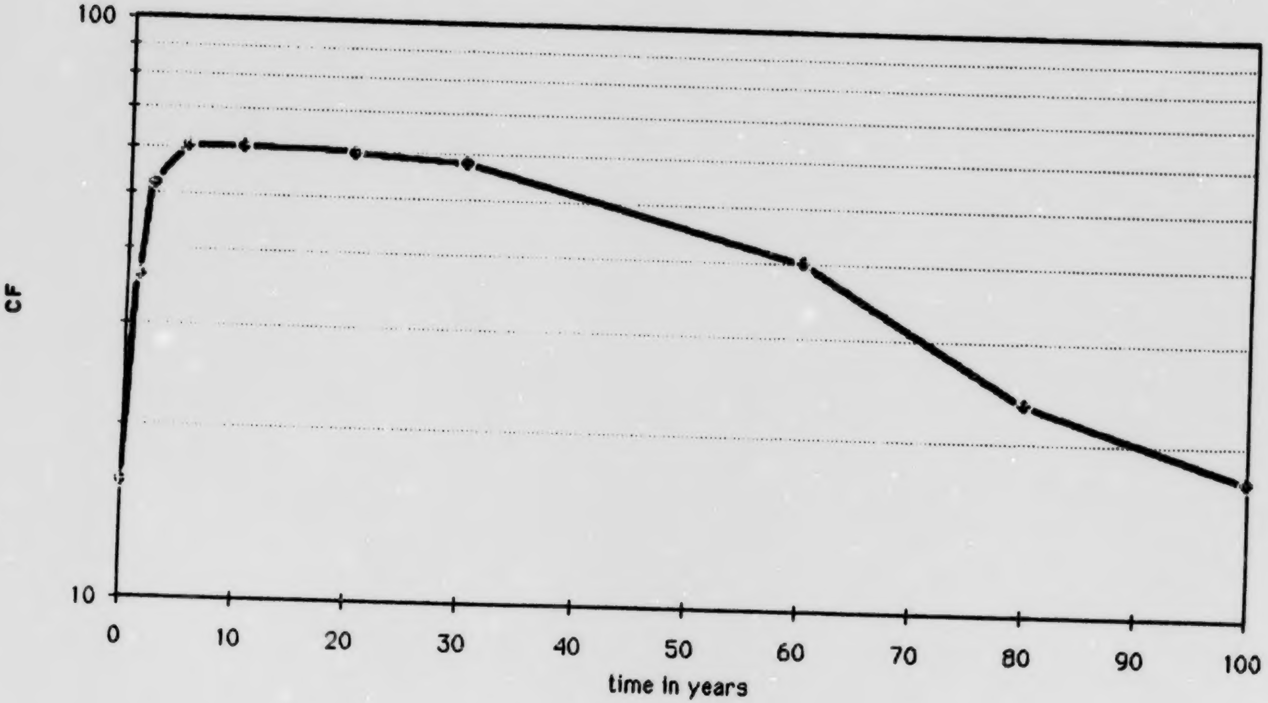
* Sensitivities, S, are calculated as the percent change in the model response due to a 1% decrease in a particular rate constant. The uncertainty index, U, is the square of the simple correlation coefficient of the log transformed model response with the log transformed values of the rate constant of interest calculated from 100 Monte Carlo simulations with all parameters varied as shown in Table 2.4. Values of S<0.1 and U<0.04 are not shown.

** The abbreviations for parameter names:

Sed - Sediment
Inv - Invertebrates
Av - Aquatic Vegetation
Forg - Forage Fish
Game - Game Fish

FIGURE 3·4

CHANGE WITH TIME
OF THE WATER-TO-GAME FISH CONENTRATION FACTOR (CF),
AFTER 99·9% OF THE RADIUM IN THE INPUT IS REMOVED



uncertainties associated with the uptake of radium by game fish flesh, the importance of the water-to-game rate constant grows from 16% in year 1 to 30% in year 10. The rate constant affecting the loss of radium from game fish accounts for an additional 23% of the uncertainties in year 10. Thus, across time and scenarios, the rate constants affecting the uncertainty of radium in game fish flesh are those associated with the direct uptake and loss of radium between water and game fish.

The geometric means and geometric standard deviations for model components of Scenario 2 are given in Table 3.4. The water component has the smallest uncertainty (geometric standard deviation ranging from 1.4 in year 1 to 1.87 in year 5) while invertebrates have the largest uncertainty (8.02 in year 1 to 7.76 in year 10). The uncertainty in the water-to-game fish flesh concentration factor is less than Scenario 1 (Table 3.2) reflecting the reduced importance of the water-to-game fish rate constant and the increased importance of the mean residence time of radium in game fish flesh.

4.0 DISCUSSION AND RECOMMENDATIONS

The development of a dynamic model of Ra-226 in an oligotrophic lake allows information to be represented in a coherent manner and the time dynamics of the system to be simulated. This capacity has not been previously achieved because of difficulties involved in formulating the rate constants from the scattered information available. The development of this model is, therefore, a significant improvement in our ability to understand these systems, compare possible scenarios, and develop future modelling and sampling experiments.

Two caveats must be kept in mind while interpreting these results: (1) the results will change as actual measurements for

Table 3.4

GEOMETRIC MEANS AND GEOMETRIC STANDARD DEVIATIONS OF MODEL PREDICTIONS FOR SCENARIO 2

	Scenario 2: 99.9 % Cleanup Case					
	Year 1		Year 5		Year 10	
Compartment	Mean	S.D.	Mean	S.D.	Mean	S.D.
Water*	6.16 E 9	1.40	1.94 E 9	1.87	1.24 E 9	1.61
Game Fish*	8.26 E 3	1.24	6.01 E 3	1.66	3.69 E 3	2.07
Forage Fish*	1.63 E 5	3.60	6.62 E 4	5.99	4.19 E 4	6.94
Aquatic Vegetation*	1.40 E 7	4.67	4.65 E 6	4.68	2.97 E 6	4.40
Invertebrates*	1.03 E 7	8.02	3.57 E 6	7.80	2.27 E 6	7.76
Sediments*	5.38 E 10	1.09	3.84 E 10	1.41	2.48 E 10	1.94
CF: Water-to-Game Fish**	4.29 E 1	1.44	9.93 E 1	2.02	9.54 E 1	2.19

* The geometric mean values equal the mass of radium-226 in each model component in pCi.

** The concentration factor (CF) equals:
[pCi Ra-226/kg fresh fish flesh] ÷ [pCi Ra-226/L water]

critical parameters become available (e.g. site specific information for the mean life expectancy of game fish). Assumptions and choices made during the development of this model have significantly affected the results (see Table 3.1 and 3.3). Different dynamics may become apparent as new experiments are performed and additional data becomes available. (2) It is important in any assessment activity to recognize the actual scenario for which an assessment is being performed. Because sensitivities and uncertainties vary with time and scenario (Tables 3.1 and 3.3), a misspecification of the assessment scenario may result in errors greater than those due to uncertainties in the rate constants or errors in model formulation.

Keeping the above caveats in mind, the model and simulations presented here lead to several recommendations. The use of a deterministic water-to-game fish flesh concentration factor (CF) based on measurements of the system at equilibrium can be misleading. For instance, in extreme cleanup situations (e.g. Scenario 2) the deterministic equilibrium CF of 16.0 underpredicts the actual concentration in game fish flesh by a factor of 4. This problem may be resolved by precise information on the residence time of radium in game fish; however, these measurements are difficult to obtain under field conditions and must be made on a site-specific basis. In a clean lake which is being contaminated, the deterministic equilibrium CF overpredicts the concentration of radium in flesh of game fish for the first 10 years. Nevertheless, all of the simulated CF values are within the range of observed values of CF reported in the literature.

We emphasize that the uncertainties estimated for predictions of this model are a function of the uncertainties associated with the rate constants but not of the uncertainties associated with the estimation of the biomass of the aquatic organisms. If the uncertainties of the biomass of game fish (i.e. the geometric standard deviation) are as little as 1.6 (the actual value is likely to be much larger), then the uncertainty in the game fish flesh CF predicted by the model will be at least as large as the measured values of the CF. Inclusion in the model of measured uncertainties in game fish biomass should resolve this problem.

The complexity of this model is desirable for studying and understanding the dynamics of radium in the aquatic system, but a much simpler system may be used for making reliable predictions. For instance, it is evident from the cases run with the model that over 90% of the radium activity in game fish flesh comes directly from the water. This means that the aquatic food chain is relatively unimportant. The most important factors are the processes governing the concentration of radium in the flesh of fish, the mean life expectancy of the game fish and the water to game fish transfer dynamics. A reduced model with three components (water, sediments and game fish) and three parameters (biomass, water-to-game transfer, and the mean residence time) would adequately capture the important dynamics of the more complexing system.

Finally, it is important to note, from the standpoint of health physics, that the ingestion of Ra-226 in the flesh of game fish may not be as significant to the exposure of critical population groups as is the direct consumption of lake water. Even if a maximum CF of 200 is assumed, more than 80 grams of fish would have to be consumed per week before the radium intake by a fisherman would exceed the intake of a person drinking 2.2 liters of lake water per day. A maximum CF of 200 would be equivalent to the upper bound of the lognormal distribution of reported values of CF, whereby the upper bound is defined as the product of the square of the geometric standard deviation and the geometric mean.

REFERENCES

Anderson, J.B., Tsivoglou, E.C. and Shearer, S.D., 1963. "Effects of Uranium Mill Wastes on Biological Fauna of the Animas River (Colorado-New Mexico). In Radioecology, Proceedings of the First National Symposium on Radioecology. Rheinhold Publishing Corp. New York, pp. 373-383.

Blaylock, B.G., 1982. "Radionuclide Data Bases Available for Bioaccumulation Factors for Freshwater Biota." Nuclear Safety, Vol. 23 (4), pp. 427-437.

Forsythe, G.E., Malcolm, M.A., Moler, C.B., 1977. "Computer Methods for Mathematical Computation." Prentice-Hall Inc.

Gardner, R.H., 1984. "A Unified Approach to Sensitivity and Uncertainty Analysis." In: Applied Simulation and Modelling, M.H. Hamza, (ed.), Proceedings of the IASTED International Symposium, San Francisco, U.S.A., June 4-6, 1984. ACTA Press, Anaheim, Ca., pp. 155-157.

Halbert, B.E., 1985. Personal Communication. SENES Consultants Limited, Willowdale, Ontario.

Halbert, B.E., Stone, D., and Cassaday, V.J., 1985. "Discussion Paper on the Data Base Available for Estimating Radium-226 Bioaccumulation in Fish Flesh." Prepared for the National Uranium Tailings Program, CANMET, EMR, Ottawa (in this report).

Hesslein, R.H. and Slavicek, E., 1984. "Geochemical Pathways and Biological Uptake in Small Canadian Shield Lakes." Can. J. Fish. Aquatic Sci., Vol. 41, pp. 459-468.

Hoffman, F.O. and Gardner, R.H., 1983. "Evaluation of Uncertainties in Environmental Radiological Assessment Models." In Radiological Assessment: A Textbook on Environmental Dose Assessment, J.E. Till and H.R. Meyer (eds.), U.S. Nuclear Regulatory Commission, Washington, D.C. NUREG/CR-3332, ORNL-5968, pp. 11-1 to 11-55.

Hoffman, F.O., Gardner, R.H. and Blaylock, B.G., 1985. "The Transfer of Radium from the Atmosphere to Pasture Vegetation." Prepared for The National Uranium Tailings Program, CANMET, EMR, Ottawa (in this report).

Kimmel, B.L. and Soballe, D.M., 1985. Personal Communication. Oak Ridge National Laboratory, Oak Ridge, TN.

National Council on Radiation Protection and Measurements, 1984. "Radiological Assessment: Predicting the Transport, Bioaccumulation, and Uptake by Man of Radionuclides Released to the Environment." NCRP Report No. 76, Bethesda, MD.

Nelson, D.J., 1966. "The Prediction of Sr-90 Uptake in Fish Using Data on Specific Activities and Biological Half Lives." In Radioecological Concentration Processes. Pergamon Press, Oxford and New York, pp. 843-851.

Ontario Water Resources Commission, 1972. "Water Pollution from the Uranium Mining Industry in the Elliot Lake and Bancroft Areas: Vol. 2 - Detailed Findings." Toronto, Ontario.

Ophel, I.L., 1978. "Aquatic Food Chain Transport of Radionuclides." In Proceedings of Workshop on the Evaluation of Models Used for the Environmental Assessment of Radionuclide Releases." CONF-770901, Oak Ridge National Laboratory, Oak Ridge, TN.

Rope, S.K. and Whicker, F.W., 1980. "Radium Bioaccumulation in Trout, Final Report for the Period Dec. 1, 1978 - Nov. 30, 1980." Department of Radiobiology and Radiation Biology, Colorado State University, Fort Collins, Colorado.

Rope, S.K. and Whicker, F.W., 1985. "A Field Study of Radium Accumulation in Trout with Assessment of Radiation Dose to Man." Health Physics, (in press).

Rubinstein, R.Y., 1981. "Simulation and the Monte Carlo Method." John Wiley and Sons.

Scharer, J.M. and Halbert, B.E., 1985. "Discussion Paper on Modelling Radium Exchange Between the Water Column and Lake Sediment." The National Uranium Tailings Program, CANMET, EMR, Ontario (in this report).

Schindler, D.W., and Holmgren, S.K., 1971. "Primary Production and Phytoplankton in the Experimental Lake Area, Northwestern Ontario, and other Low-carbonate Waters, and Liquid Scintillation Method for Determining C-14 Activity in Photosynthesis." Journal Fisheries Research Board of Canada, Vol. 23, pp. 189-201.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long Term Effects of Uranium Mill Tailings in Canada - Phase I." Research Report Prepared for the National Uranium Tailings Program, CANMET, EMR, Ottawa, Under Supply and Services Canada Contract Number OSQ83-00236.

Swanson, S.M., 1983. "Levels of Ra-226, Pb-210 and Total U in Fish Near a Saskatchewan Uranium Mine and Mill." Health Physics, Vol. 45(1), pp. 67-80.

Swanson, S.M., 1985a. Personal Communication. Saskatchewan Research Council, Saskatoon, Saskatchewan.

Swanson, S.M., 1985b. "A Study of the Distribution of Naturally Occurring Radionuclides in Freshwater Benthos and Their Environment: Phase II Report." Report by SENES Consultants Limited and the Saskatchewan Research Council for the National Uranium Tailings Program, CANMET, EMR, Ottawa. Under Supply and Services Canada Contract Number OSQ83-00241.

Wetzel, R.G., 1975. "Limnology." W.B. Saunders Co., Toronto.

DISCUSSION PAPER ON

THE DATA BASE AVAILABLE FOR ESTIMATING RADIUM-226 BIOACCUMULATION IN FISH FLESH

by
B.E. Halbert, D. Stone and V.J. Cassaday
SENES Consultants Limited
Willowdale, Ontario

June, 1985

Abstract

Many processes may contribute to Ra-226 levels observed in fish flesh. The bioaccumulation factor (BF) is an aggregate, single-value parameter which is often used to represent the effect of these processes in modelling the transfer of Ra-226 from water to fish flesh. This paper summarizes values of the BF's for Ra-226 compiled from published studies undertaken at uranium mining locales in Canada. These data are highly variable and strongly skewed toward lower values suggesting a lognormal behaviour. Recommended statistics for Ra-226 transfer from water to fish flesh are presented.

1.0 INTRODUCTION

Pathways analysis provides the link between the release of radioactive or non-radioactive contaminants to the environment and the exposure of man to these contaminants. In pathways analysis, the passage of a contaminant through the environment is traced along the following sequence of steps: release to the environment, transport through the air and water, uptake by biota, and human exposure.

Modelling the uptake of radionuclides by aquatic biota can be quite involved and requires consideration of several interactions (for example, aquatic plants decomposing and returning radionuclides to the sediment or water phases, predatory fish eating smaller fish, etc). In a separate investigation into the applicability of a multi-component model for predicting the dynamics of radium-226 (Ra-226) transfer to game fish flesh, twenty three (23) pathways were identified (Gardner et al 1985). One of the conclusions drawn from that investigation was that a reduced model would adequately capture the important dynamics of the more complex system. Indeed, the approach being taken in the mainstream system modelling is to aggregate these processes and represent them by a single transfer coefficient, namely, the bioaccumulation factor (BF) or concentration factor (CF). It is important to keep in mind however, that the aggregate model implicitly includes all of the other processes.

The use of a single BF value assumes that a quasi-steady state condition prevails. Such a condition implies a relatively constant relationship in the contaminant concentrations between the system components and hence, the value of the BF will be reasonably constant (Whicker and Schultz 1982). Even assuming that a quasi-steady state condition does exist, other variables can influence the selection of the BF. These include the chemical form of the radionuclide being considered, the possible range in the constituent level in the aqueous phase, the type of fish (i.e. predator vs. scavenger) and the chemical, physical and biological attributes of the system.

This paper presents a summary of available data on the concentration of Ra-226 in fish flesh relative to the concentration measured in water at the same time and location. The emphasis in this review has been placed on compiling BF's for fish flesh as this is the primary pathway of dose to man from consumption of contaminated aquatic biota. A brief review of transfer coefficients reported in the literature is also presented for comparison purposes.

2.0 FACTORS AFFECTING RADIUM DISTRIBUTION

Radium belongs to group IIA in the periodic table, the alkaline earth elements. It has chemical properties similar to those of barium, has one oxidation state (+2) in solution, is highly basic and not easily complexed (Sheppard 1980). Ra-226 has the longest half-life of the radium isotopes and is the one most frequently studied with respect to food-chain transfer.

Ra-226 is relatively insoluble with environmental levels typically reported in the pg/L range. Upon entering the water column, it reportedly moves quickly to sediments through adsorption or co-precipitation with barium or calcium (Sheppard 1980, Dyck 1974, Emerson and Hesslein 1973). Morin et al (1981) claim that removal of Ra-226 to sediments is controlled by adsorption more than co-precipitation.

The three major parameters controlling adsorption onto sediments are particle size, organic content (whether or not it is part of living organisms) and manganese or iron content. Adsorption increases with decreasing particle size, and is greatest at sizes less than 100 um (Sheppard 1980, Dave et al 1981). Adsorption is usually greater in sediments with high organic content (Dyck 1974, Titayeva et al 1978). Finally, hydrous oxides of iron and manganese have long been known to adsorb trace elements from solution.

Once deposited onto sediments, Ra-226 can be re-distributed by physical, chemical and biological phenomena. High-flow periods in streams can scour organic material off the bottom and move finer sediments further downstream. Spring and fall turnover in lakes can transport littoral sediment material to profundal zones. Movement of Ra-226 from littoral to profundal areas was observed in the artificially-injected lakes studied by Emerson and Hesslein (1973). Sorption-desorption phenomena can also result in Ra-226 redistribution in sediments. For example, in anoxic waters in the hypolimnion of stratified lakes, manganese is reduced, releasing Ra-226 back into solution. When turnover occurs, the Ra-226 is re-adsorbed and redistributed (Hesslein and Slavicek 1984). Desorption from sediments is also related to the relative concentration of Ra-226 in water and sediment. Desorption increases as relative water concentration decreases via diffusion from sediment porewater (Hesslein 1980). Cycling of Ra-226 through food chains also results in redistribution, as does the burrowing activity of sediment-dwelling invertebrates.

The general ecological properties of Ra-226 include: a chemical analogue of calcium; primary exposure routes are via ingestion, absorption or external gamma irradiation; a moderate degree of food chain transport; a successive trophic level concentration of less than 1.0; accumulation in bone; a moderate degree of assimilation; and a long (in the order of years) biological half-life (Whicker and Schultz 1982).

One of the most important ecological properties governing the distribution of Ra-226 in ecosystems is its similarity to calcium, an essential nutrient for many biota. Ra-226 follows the same biological pathways as calcium, but is favoured or discriminated against to a greater or lesser extent in different processes and organisms (Hesslein and Slavicek 1984).

Because Ra-226 is an analogue of calcium, it is primarily deposited in hard tissues such as bone or exoskeleton (thus the term "bone-seeking"). Bone-seeking elements are generally absorbed to a smaller extent than those that concentrate in soft tissue, but are retained more tenaciously (Whicker and Schultz 1982). Because they are deposited in relatively undigestible tissues (metabolic sinks), their mobility in food chains is only moderate (Whicker and Schultz 1982, Reichle et al 1970).

3.0 RADIUM TRANSFER TO FISH FLESH

Most reviewers agree that the data base for describing the behaviour of Ra-226 in aquatic freshwater systems is limited. That data which are available comes mostly from studies of the effects of uranium mining and milling operations. Table 1 summarizes a literature survey of available (directly available and derived) BF's conducted for a recent environmental assessment (SENES 1983). The data reported in the table indicate a fairly wide range in possible values. For comparison purposes, the data reported by Swanson (1982) for both fish flesh and whole fish samples have been included. These data demonstrate the tendency of Ra-226 to deposit primarily in hard tissue, such as bone or exoskeleton, as previously discussed.

To augment the above data, the literature and, in particular, recently-prepared environmental assessment documents on uranium operations in Canada, were reviewed to identify additional sources of data of Ra-226 levels in fish flesh and water samples taken at comparable times and locations. Once the data were compiled and culled to remove unusable data (i.e. all results reported as less than values were deleted as were results for which the basis of measurement - eg. dry or wet weight - were not reported), the data base was segregated according to whether the fish species could be classified as game fish or forage fish. The interest in this separation in the data base was to determine whether or not there was any measurable difference in the BF for forage fish which presumably consume more contaminated material. Similar BF values for the two groups of fish would suggest that the main source of Ra-226 was from exposure to the water phase and not from the food consumed.

The data bases for analyzing BF values for game fish and forage fish are summarized in Table 2 and 3 respectively. In total, 44 game fish and 35 forage fish BF values are given on the tables. The summary statistics on both data bases suggest that the data are quite variable. BF values for Ra-226 in game fish ranged from 1 to 603 $L.kg^{-1}$ as compared to BF values for forage fish which varied from 3 to 90 $L.kg^{-1}$. Examination of the two data sets indicates that most of the values are located near the low end of the ranges which suggests that frequency plots would be highly skewed to the left with long tails to the right. The geometric statistics therefore, are more meaningful.

The geometric mean values for the two data bases are quite comparable. The mean fish flesh BF value for game fish equals 15.9 $L.kg^{-1}$ compared to a value of 12.8 $L.kg^{-1}$ for forage fish. This comparison suggests that the premise that the principal pathway of Ra-226 transfer to fish flesh is from water is correct. The geometric mean values given in Tables 2 and 3 are also in reasonable agreement with the literature cited BF values summarized in Table 1.

The geometric standard deviation for game fish of 3.6 is quite high, again reflecting the highly variable (and uncertain) nature of the data base. To improve the predictions in a probabilistic analysis, it is evident that site-specific data are needed. For the purposes of the generic assessment, however, the summary statistics given on Table 2, assuming a lognormally distributed data base, are recommended.

Table 1

Review of Literature-Reported BF Values for Radium-226 in Fish Flesh

Reference	Bioaccumulation Factor[1]	Comments
Eldorado 1981	11	Beaverlodge area only.
Elliot et al 1981	20.2 (12.9-120)	Median (range) (n=13).
Canada Wide Mines 1981	15	Derived from data referenced in the report.
Swanson 1982	28 (5-80)	Average (range) (n=5 locations) Large fish species (flesh).
Swanson 1982	593 (67-833)	Average (range) (n=5 locations) Forage fish (whole).
Blaylock 1982[2]	0.3-137	Literature review.
U.S. NRC, 1976	50	Regulatory guide.
Zach 1980	50	NEPTUN computer model

Notes

1. All bioaccumulation factors are expressed as the ratio of the radium content of fish to the radium concentration in water and are presented in units of $L.kg^{-1}$ fresh weight (i.e. $Bq.kg^{-1}.Bq^{-1}.L = L.kg^{-1}$).

2. Blaylock also quotes a study by Ophel, in which the effect of mineral content in fresh waters on transfer coefficients was examined. The coefficient for radium in water of low mineral content was 150; high mineral content was 10.

Table 2

Game Fish Flesh Ra-226 Bioaccumulation Factors

Location	Species	No. Fish Samples	Tissue Analyzed	Fish Flesh Concentration pCi/kg (fresh)	Water Level pCi/L	BF (L/kg)	Reference
B.C. Blizzard Site							
Sandrift Lake #2	Rainbow Trout	5	flesh	71	0.2	355	Norcen 1980
Beaverdell Creek	Rainbow Trout	14	flesh	3	0.2	15	
Saskatchewan							
Kewen Lake	Walleye	1	flesh	4	0.1	40	Canadian Occidental 1981
	"	1	"	13.8	0.1	138	
	"	1	"	8.4	0.1	84	
	"	1	"	4.3	0.1	43	
	"	1	"	60.3	0.1	603	
McClean Lake	Walleye	1	flesh	2.4	0.1	24	Canadian Occidental 1981
	"	1	"	27	0.1	270	
	"	1	"	2.4	0.1	24	
	"	1	"	13.3	0.1	133	
	"	1	"	14.9	0.1	149	
Beaverlodge Lake	Lake trout	9	flesh	5.44	1.5	3.6	Swanson 1980
	" "	10	flesh	–	–	1	Swanson 1983
Key Lake Area							I.E.S. Inc. 1983
McDonald Lake	Northern Pike	5	flesh	.01	–	10	
Delta Lake	Northern Pike	5	"	–	–	10	
Russell Lake	Northern Pike	5	"	–	–	5	
Dubyna Lake	Lake Trout	1	flesh	3.4	0.5	6.8	SRC 1978
	" "	1	"	3.4	0.5	6.8	
	" "	1	"	3.4	0.5	6.8	
	" "	1	"	6.8	0.5	13.6	
	Northern Pike	1	flesh	13.2	0.5	26.4	
	" "	1	"	13.2	0.5	26.4	
	" "	1	"	21.4	0.5	42.8	
	" "	1	"	13.2	0.5	26.4	
	" "	1	"	13.2	0.5	26.4	
Foot Bay of Donaldson Lake	Lake Trout	1	flesh	3.4	0.9	3.8	SRC 1978
	" "	1	"	3.4	0.9	3.8	
	" "	1	"	3.4	0.9	3.8	
	" "	1	"	6.8	0.9	7.6	

Table 2 continued

Location	Species	No. Fish Samples	Tissue Analyzed	Fish Flesh Concentration pCi/kg (fresh)	Water Level pCi/L	BF (L/kg)	Reference
Schmoo Lake	Lake Trout	1	flesh	6.8	.535	12.7	SRC 1978
	" "	1	"	3.4	.535	6.4	
	" "	1	"	6.8	.535	12.7	
	" "	1	"	3.4	.535	6.4	
	" "	1	"	3.4	.535	6.4	
Wollaston Lake							
Pow Bay	Lake Trout	5	flesh	0.3	.008	37.5	ELDOR (Gulf Minerals) 1982
	Northern Pike	5	flesh	0.2	.008	25	
Hidden Bay	Northern Pike	5	flesh	0.2	.014	14.2	
Key Lake Area	Walleye	1	flesh	.9	.16	5.6	Key Lake 1979
	"	1	"	.99		6.2	
	"	1	"	2.0		12.5	
	"	1	"	.86		5.3	
	"	1	"	1.1		6.9	
	"	1	"	1.0		6.2	

SUMMARY STATISTICS FOR BIOACCUMULATION FACTOR (BF)

Number of Samples: n = 44
Mean Values ($\bar{X}$): Arith $\bar{X}$ = 44.4 Geom $\bar{X}$ = 15.9
Standard Deviations (σ) Arith σ = 100.0 Geom σ = 3.6
Range of BF Values: 1 to 603

Table 3

Forage Fish Flesh Ra-226 Bioaccumulation Factors

Location	Species	No. Fish Samples	Tissue Analyzed	Fish Flesh Concentration pCi/kg (fresh)	Water Level pCi/L	BF (L/kg)	Reference
B.C. Blizzard Site Trapper Lake	Sucker	22	flesh	2	0.2	10	Norcen 1980
B.C. Atlin Area Ruby Creek	Arctic Grayling	1	composite (whole small fish pooled)	7	0.5	14	Placer Development 1981
Boulder Creek	" "	1		4	0.5	8	
Pine Creek	" "	1		5	0.4	12.5	
Saskatchewan Kewen Lake	Whitefish	1	flesh	2	0.1	20	Canadian Occidental 1981
	"	1	"	9	0.1	90	
	"	1	"	6	0.1	60	
	"	1	"	2	0.1	20	
	"	1	"	3.9	0.1	39	
McClean Lake	Whitefish	1	flesh	5.48	0.1	54.8	Canadian Occidental 1981
	"	1	"	2.66	0.1	26.6	
	"	1	"	4.96	0.1	49.6	
Beaverlodge Lake	Lake Whitefish	23	flesh	-	-	3	Swanson 1983
	Longnose Sucker	1	"	-	-	12	
	Lake Whitefish	18	flesh	-	-	5	Swanson 1982
	White Suckers	17	"	-	-	18	
Key Lake Area McDonald Lake	Lake Whitefish	5	flesh	-	-	10	I.E.S. Inc. 1983
Delta Lake	Lake Whitefish	5	"	-	-	30	
Russell Lake	Lake Whitefish	5	"	-	-	6	
Russell Lake	Lake Whitefish	1	flesh	1.1	.16	6.8	Key Lake 1979
	" "	1	"	1.2	.16	7.5	
	" "	1	"	1.1	.16	6.8	
	" "	1	"	1.2	.16	7.5	
	" "	1	"	1.0	.16	6.2	
	" "	1	"	1.1	.16	6.9	
Russell Lake	Longnose Sucker	1	flesh	1.2	.16	7.5	Key Lake 1979
	" "	1	"	2.1	.16	13.1	

Table 3 continued

Location	Species	No. Fish Samples	Tissue Analyzed	Fish Flesh Concentration pCi/kg (fresh)	Water Level pCi/L	BF (L/kg)	Reference
Russell Lake cont'd	Longnose Sucker	1	flesh	1.3	.16	8.1	Key Lake 1979
	" "	1	"	1.2	.16	7.5	
	" "	1	"	1.1	.16	6.9	
Beaverlodge Lake	Lake Whitefish	22	flesh	.2	.06	3.3	Swanson 1984
	White Suckers	20	"	1.0	.06	16.7	
Milliken Lake	Lake Whitefish	24	"	.2	.01	20	Swanson 1984
	White Sucker	18	"	.1	.01	10	
Fredette Lake	Lake Whitefish	21	"	.4	.02	20	Swanson 1984

SUMMARY STATISTICS FOR BIOACCUMULATION FACTOR (BF)

Number of Samples: $n = 35$
Mean Values ($\bar{X}$): Arith $\bar{X} = 18.4$ Geom $\bar{X} = 12.8$
Standard Deviations (σ) Arith $\sigma = 19.0$ Geom $\sigma = 2.3$
Range of BF Values: 3 to 90

DISCUSSION PAPER ON

THE DATA BASE AVAILABLE TO DEVELOP PARAMETER DISTRIBUTIONS FOR USE IN MODELLING THE TRANSFER OF AIRBORNE RADIOACTIVITY TO TERRESTRIAL VEGETATION

Prepared by

V.J. Cassaday, B.E. Halbert and D. Stone
SENES Consultants Limited
Willowdale, Ontario
June 1985

Abstract

The application of probabilistic assessment concepts to modelling the behaviour and effects of uranium mill tailings requires that the various pathways of contaminant transport from the source to man be assessed. This paper examines the data base available for characterizing parameters included in models of terrestrial transport of radionuclides. Published data on both direct foliar retention and root uptake mechanisms are reviewed for five radionuclides; radium-226, natural uranium, lead-210, polonium-210 and thorium-230. Where possible, the data are segregated into one of four vegetation groupings: grass, hay and herbage; above-ground vegetables; below-ground vegetables; and, fruit.

TABLE OF CONTENTS

Page

1.0 INTRODUCTION 1

2.0 TERRESTRIAL VEGETATION PATHWAYS MODELLING APPROACH 1
- 2.1 Background 1
- 2.2 Atmosphere-to-Vegetation Transfer Model ($P_{1,5}$) 2
- 2.3 Root Uptake-to-Vegetation Transfer Model ($P_{4,5}$) 5

3.0 ATMOSPHERE-TO-VEGETATION MODEL PARAMETERS 6

4.0 SOIL-TO-VEGETATION TRANSFER FACTORS 7
- 4.1 Introduction 7
- 4.2 Radium-226 Transfer Factor 9
  - 4.2.1 General 9
  - 4.2.2 Radium-226 from Soil to Grass, Hay and Herbage 10
  - 4.2.3 Radium-226 from Soil to Vegetables (Above Ground) 11
  - 4.2.4 Radium-226 from Soil to Root and Stem Vegetables 11
  - 4.2.5 Radium-226 from Soil to Fruit 12
- 4.3 Uranium Transfer Factors 13
  - 4.3.1 General 13
  - 4.3.2 Uranium from Soil to Grass, Hay and Herbage 13

Table of Contents (continued) Page

4.3.3 Uranium from Soil to Vegetables (Above Ground) 14
4.3.4 Uranium from Soil to Root and Stem Vegetables 14
4.3.5 Uranium from Soil to Fruit 15
4.4 Lead-210, Polonium-210 and Thorium-230 15
Transfer Factors 17

REFERENCES

1.0 INTRODUCTION

Probabilistic analysis requires that appropriate frequency distributions be ascribed to the model parameters. The focus of this paper is on a review of the data base available for radionuclide transfer to vegetation. To facilitate the discussion a review of the models used to described transfer to vegetation is also presented.

2.0 TERRESTRIAL VEGETATION PATHWAYS MODELLING APPROACH

2.1 Background

Airborne materials can be transferred to the edible portions of vegetables consumed by man or to forage (pasture grass) consumed by domestic or wild animals via two mechanisms: direct foliar retention and root uptake. The transfer model used for the Phase I probabilistic assessment (SENES 1984) considered only the direct foliar retention of particulates. This mechanism was chosen because past experience in impact assessments for uranium developments has generally indicated that root uptake contributed only a relatively small amount to the total vegetation concentration. This conclusion is based on certain specific scenarios under consideration and it is emphasized that applications could be envisaged where root uptake could be the major route. For example, if the disposal scenario involved tailings placement at the bottom of a deep lake, then the only pathway to vegetation might be root uptake where lake water was periodically applied to vegetation via flooding.

Some investigators have questioned whether current models for calculating air deposition onto plants implicitly include root uptake (Miller and Hoffman 1983). The relative contributions of the processes of radionuclide transfer to terrestrial vegetation are being investigated separately and is the subject of a separate discussion paper. Hence only brief reviews of the model for direct deposition used in Phase I and the (conventional) model for root uptake are presented below.

2.2 Atmosphere-to-Vegetation Transfer Model ($P_{1,5}$)

For Phase I, four categories of vegetation were considered: lichen (winter forage), summer forage, above ground vegetables and below ground vegetables. Because lichens are distinctly different from the other types of vegetation they are not included in this review.

The following equation was used in Phase I to model direct foliar retention (U.S. NRC 1982):

$$C_{vi} = D_i F_r E_v \left[\frac{1 - \exp(-\lambda_u T_v)}{Y_v \lambda_w} \right]$$

The computer ease form of this equation as used in Phase I was:

$$P_{1,5} = \text{VELDEP*FDEPVE*ESUBV*} \ \frac{\text{(1-exp(-LAMWTH*TEXPDP))}}{\text{YDENST*LAMWTH}}$$

where:

D_i = VELDEP = the deposition velocity ($m.s^{-1}$)

F_r = FDEPVE = the fraction of the total deposition retained on the plant surface (0 to 1)

E_v = ESUBV = the fraction of the foliar deposition retained on edible portions of vegetation v (0 to 1)

λ_w = LAMWTH = a decay constant accounting for weathering losses (s^{-1})

T_v = TEXPDP = the duration of plant exposure for vegetation type v to atmospheric deposition (s)

Y_v = YDENST = the yield density for vegetation type v (g (wet wt).m^{-2})

C_{vi} = $P_{1,5}$ = the atmosphere-to-vegetation transfer factor ($m^3.g^{-1}$)

None of these values was given a distribution during Phase I. The deposition velocity, VELDER, depends in a complex fashion on various physical and chemical processes and was not included in this review which focusses on the biological processes.

FDEPVE is the fraction of the total deposition per unit area retained on the plant surface. It is also sometimes referred to as the interception fraction. The parameter is vegetation specific and is influenced by such factors as plant spacing and leaf type and surface area.

ESUBV is the fraction of the foliar deposition retained on edible portions of the vegetation. This parameter simply allows the user to specify which portion of the plant is to be used for the pathways analysis. It is vegetation specific, for instance the edible portion of beans, which grow under a canopy of leaves, would be expected to receive a smaller percentage of the total deposition on the plant than the edible portion of leaf lettuce, although these plants might have similar values for FDEPV. ESUBV is also used to characterize chemical/biological processes affecting radionuclide transfer within the plant tissue. For instance, root vegetables are usually assumed to receive a portion of the particulate deposited on the above ground part of the plant. For this reason, ESUBV is sometimes referred to as the translocation factor.

LAMWTH, the decay constant accounting for weathering losses, is usually combined with the radioactive decay constant to derive an effective decay constant (λ_{eff}) as follows:

$$\lambda_{eff} = \lambda_w + \lambda_r$$

where

$$\lambda_w = LAMWTH = \frac{\ln 2}{T_w}$$

T_w = the time necessary for one-half of the initial contamination deposited on a plant to be removed by environmental processes (environmental half-life)

and

$$\lambda_r = \frac{\ln 2}{T_r}$$

T_r = the radiological half-life

The parameter λ_r was not used in Phase I because for radionuclides with a long half-life (T_r) relative to the environmental half-life (T_w), the effective decay constant ($_{eff}$) is controlled by λ_w. If radionuclides with a smaller half-life are modelled (i.e. polonium-210), this term should be included in the transfer model.

The environmental half-life (T_w) is especially important when uptake rate of radionuclides is small compared to the period between initial deposition and harvest of vegetation (Miller and Hoffman 1983).

A number of physical processes affect the concentration of deposited radionuclides on vegetation with time: wind removal (particularly for particles > 1 um); water removal via wash-off and/or leaching; growth dilution, which is particularly important before herbaceous plant maturity; grazing, which acts as a removal mechanism; and stage of plant growth, which may affect any of the above processes (Miller and Hoffman 1983).

Thus factors affecting the variability of T_w are related to the physiochemical form of the radionuclide particulate, vegetation type and growth form, climate and season. The reader is referred to Miller and Hoffman (1983) for more detail.

The parameter TEXPDP, the duration of plant exposure to atmospheric deposition, is perhaps the easiest term to quantify in this model since herbaceous vegetation is being modelled. The parameter must be specified for each vegetation type under consideration.

YDENST, the yield density, is also vegetation specific.

It should be noted that this model does not allow for the decay of the radionuclide between the period of harvest and consumption, or for ingrowth from a parent.

A review of values used for the above parameters in recent assessments is presented in Chapter 3.

2.3 Root Uptake to Vegetation Transfer Model ($P_{4,5}$)

The concentration of a radionuclide in vegetation attributable to root uptake is conventionally modelled as follows (U.S. NRC 1982):

$$C_{vi} = C_{gi} \frac{B_{vi}}{p}$$

where

C_{gi} =the concentration of radionuclide "i" on the soil surface ($Bq.m^{-2}$)

C_{vi} =the resulting concentration of radionuclide "i" in vegetation "v" ($Bq.g^{-1}$)

B_{vi} =the soil-to-plant transfer factor for radionuclide "i" and vegetation type "v" (Bq.g^{-1} (wet) plant per Bq.g^{-1} (dry) soil)

p = the assumed soil areal density for surface mixing, 240 kg (dry weight).m^{-2}

For the multiplicative chain model defined in the Phase I study (SENES 1984), the transfer model equates to:

$$P_{4,5} = \frac{B_{vi}}{p}$$

where

$P_{4,5}$ = the root uptake-to-vegetation transfer factor is equal to $C_{vi}.C_{gi}^{-1}$ (m^2.g^{-1} (wet wt))

Plant uptake of radionuclides from the soil is affected by numerous processes and factors. As a result, B_{vi} may vary widely. Some of the influencing factors include: the physiochemical form of the radionuclide; type of vegetation; soil characteristics (including stable element concentrations and the presence of fertilizers, agricultural chemicals and chelating agents); and the distribution of radionuclides within the soil (Peterson 1983). A review of transfer factors reported in the literature for those radionuclides of interest to the Phase II investigation is presented in Chapter 4.

3.0 ATMOSPHERE-TO-VEGETATION MODEL PARAMETERS

The literature indicates that some of the parameters used as inputs in the model of direct deposition to vegetation do vary with the physical/chemical characteristics of the depositing material (i.e. LAMWTH and ESUBV (see Section 2.2)). However, no information specific to the radionuclides of interest was found in our review. Rather, information was classified by the type of vegetation.

Table 3.1 summarizes the values presented in recent publications for the types of vegetation (except lichens) modelled in Phase I.

4.0 SOIL-TO-VEGETATION TRANSFER FACTORS

4.1 Introduction

This chapter presents separate reviews of the data on radium and uranium transfer factors from soil to vegetation and their principal governing factors. Data on the transfer of thorium, lead and polonium from soil to vegetation is also discussed but because the data base is very sparse and poorly documented, these radionuclides are reviewed as a group.

Much of the literature was found to lack a common approach to defining and reporting concentrations and transfer factors. This review attempts to account for some of the variability in published transfer factors and provide selected credible values for future modelling of the effects of Canadian uranium tailings facilities.

The calculation of transfer factors by various authors has been reported on several different bases: wet plant - dry soil; dry plant - dry soil; wet plant - wet soil; and, occasionally as plant ash weight - dry soil. Some authors provide interconversion data, for example, Ng et al (1968) gave the water content of fresh vegetation as 75%, while Baker (1977 cited by Zach 1978) used 80%.

An extensive list of water contents of food and forage crops is given by Spector (1960), cited by McDowell-Boyer et al (1979). Water content of edible portions of vegetables ranges from 73.9 to 96.1 percent, fruit 74.8 to 89.3 percent, grain 8.3 to 12.5

Table 3.1

Survey of Parameter Values for Modelling Vegetation Concentration of Radionuclides Due to Air Deposition

Parameter	Grass	Above-Ground Vegetables	Below-Ground Vegetables	Reference
FDEPVE (0 to 1)	0.2	0.2	0.2	U.S.NRC 1982
	range: 0.02 to 0.82	-	-	Miller & Hoffman 1983
	$1.0^{a,b}$	$0.6^{a,b}$	$0.6^{a,b}$	Zach 1982(b)
	not included	not included	not included	CSA 1984
ESUBV (0 to 1)	1.0	1.0	0.1	U.S.NRC 1982
	$1.0^{a,b}$	$1.0^{a,b}$	$1.0^{a,b}$	Zach 1982(b)
	not included	not included	not included	CSA 1984
TEXDEP (days)	30	60	60	U.S. NRC 1982
	45^b	90^b	90^b	Zach 1982(b)
	30	60	60	CSA 1984
	geo mean 30 max[c] 200 min[c] 15	geo mean 100 max[c] 180 min[c] 60[d]	100	Hoffman et al 1984
LAMWTH expressed as T_w(d)	14	14	14	U.S. NRC 1982
	geo mean 17 range 9-34	-	-	Miller & Hoffman 1983
	geo mean 12 range 8-26	geo mean 20 range 6-63[d]	14	Hoffman et al 1984
	18	18	18	Zach 1982(b)
	14	14	14	CSA 1984
YDENST ($kg.m^{-2}$)	0.75 (wet)	2.0 (wet)	2.0 (wet)	U.S. NRC 1982
	range 0.04-1.6 (dry)	range 0.36-5.3 (fresh)	-	Miller & Hoffman 1983
	0.28 (dry)	leafy 1.9 non leafy 0.57 (fresh)	2.6 (fresh)	CSA 1984
	1.5^b (wet)	1.0^b (wet)	1.0^b (wet)	Zach 1982(b)
	0.7 (wet)	0.145-1.35 (wet)	0.357-1.30 (wet)	Boone and Ng 1981

a actually represents ESUBV x FDEPVE
b author states intentionally conservative
c 95% range
d non leafy vegetables

percent, feed grain and hay 8.8 to 15.7 percent and forage grass 80 percent.

In the following sections, grass, hay and herbage are regarded as cattle fodder and transfer factors from soil to plants are calculated on a dry weight basis. For all other vegetation, transfer factors have been calculated on, or converted to, a wet plant - dry soil basis which is the form used in calculating internal dose from food intake. The transfer factors for each radionuclide and plant category are tabulated with source references and notes on any special features of the data.

Soil-to-plant transfer factors (B_{vi}'s) are also influenced by various interrelated soil properties, including texture, clay content, dominant clay mineral, cation exchange capacity, exchangeable cations, pH and organic matter content (Nichita et al, cited by Ng 1982). Uptake varies with chemical and physical forms of the nuclide, plant species, plant part and stage of growth, as well as with experimental conditions and the manner in which the isotope is introduced into the soil. As a result, B_{vi}'s exhibit much more variability than transfer coefficients for animal products (Ng 1982). Generic transfer factors for plant-soil can vary by up to three orders of magnitude (McDowell-Boyer et al 1980).

The distributions of transfer factors reported are generally skewed in shape with a peak (modal) concentration at some relatively low level and a long tail toward the high end of the scale. Tracy et al (1983) found radionuclide concentrations in garden vegetables at Port Hope to be lognormally distributed, and used the geometric mean and it 68% confidence limits to indicate overall nuclide behaviour. Sheppard and Thibault (1983) found that the uranium content distribution of ten different plant species were significantly different from normal and that the lognormal distribution was the most appropriate for describing the distribution in most species. The data compiled in the present paper were also found to be lognormally distributed.

4.2 Radium-226 Transfer Factors

4.2.1 General

The natural distribution of radium in soils has been found to be positively correlated with organic content and absorbed bases. Most radium is strongly bound in the soil and its mobile forms in decreasing order, have been reported to be acid-soluble, exchangeable and water soluble (Taskayev cited by MacLaren 1979).

An increased supply of calcium in soils reduces the absorption of radium by a plant but the relationship between calcium and radium has not always been found to be consistent (Russel 1966 cited by MacLaren 1979). An example of the influence of soil calcium levels on the accumulation of radium by food crops is given by Lindeken (1977), cited by MacLaren (1979) as follows.

Soil Calcium Content (ppm)	Ra Transfer Factor Plant-to-Soil [pCi/g (dry wt)]/[pCi/g (dry wt)]	
	Broccoli	Turnips
3100	0.06	0.05
6200	0.02	0.03

The behaviour of radium and barium in the soil-plant system is almost identical and Taskayev (1977), cited by MacLaren (1979) showed that the behaviour of radium in a sod-meadow soil is controlled by barium rather than calcium.

Garden vegetables grown in soil contaminated with Ra-226, Pb-210 and uranium at Port Hope, Ontario, showed uptake consistent with a linear dependence on soil concentration. Highest concentrations were found in root and stem vegetables, and for radium, also in leafy vegetables. Fruit generally had the lowest values (Tracy et al 1983). Popova et al (cited by Tracy et al 1983), found radium localized mainly in roots but also noted elevated levels in leafy vegetables such as spinach and lettuce. They suggested that the radium moves quickly to areas of maximum plant growth such as root tips and terminal leaves, where it becomes permanently fixed. Mistry et al (1965) also found Ra-228 concentration highest in leaves. D'Souza and Mistry (1970) showed Ra-226 to be transferred initially from roots to shoots in bean plants.

4.2.2 Radium-226 from Soil to Grass, Hay and Herbage

In 19 examples from 8 references, the radium-226 concentration in grass, hay and herbage was shown to be lower than in the medium in which they were grown, i.e., natural soil, uranium mine and mill tailings, or experimental pot soil. Table 4.1 lists the transfer factors and data sources.

Most of the transfer factors shown are based on related soil and vegetation samples. Equilibrium conditions are assumed in all cases. The data are seen to vary over a wide range; the minimum and maximum values given on Table 4.1 are 0.008 and 0.71 respectively. The experimental results reported by Kirchmann et al (1966) are in the upper part of the observed range, and seem to reflect, for rye grass at least, differences in method. Soil surface contamination by Ra-226 seems to penetrate the plants only about 10% as much as evenly mixed contamination which surrounds the root system. The difference between first and

Table 4.1

Radium-226 Transfer Factors: Grass, Hay and Herbage from Soil

Plant species	TF*	Notes	Reference
grass, unidentified	0.135	collected from U mineralized site	Norcen 1980
pine grass	0.064		
grass, unidentified	0.133		SENES 1982
rye grass: 1st cut	0.71	soil uniformly contaminated	Kirchmann et al 1966
2nd cut	0.53		
clover: 1st cut	0.48		
2nd cut	0.32		
rye grass	0.08	soil surface contaminated	
clover	0.27		
mixed grasses	0.03	grown on U tailings	Moffett and Tellier 1977
pasture grass	0.018		U.S. NRC 1979
hay	0.082		
mixed grasses	0.21		Whicker 1978
alfalfa	0.04		
"herbage"	0.008		DeBortoli and Gaglione 1972
forage, feed, hay	0.09	unweighted average based on 89 derived values	McDowell-Boyer et al 1980
mixed grasses	0.035	natural background	Ibrahim et al 1982
mixed grasses	0.035	U tailings reclamation area	
mixed grasses	0.23	exposed U tailings	
mixed grasses	0.28	U tailings pond edge	

* Transfer factor (TF) based on radium-226 content on a dry weight plant/dry weight soil basis.

second cut transfer factor values suggests that available Ra is absorbed and transferred to early growing parts of the plants while later growth, supported by about the same root system, receives less radium.

4.2.3 Radium-226 from Soil to Vegetables (Above Ground)

Transfer factors for vegetables from soil were obtained from six main references. Because of a scarcity of data on above-ground edible portions alone, transfer factors were calculated for some foliage and pods which would not normally be eaten by people, but which appeared not unlike the more favoured portions with respect to radium concentrations. Further, some of the values used for calculating transfer factors were not from closely related vegetables and soils (Klement 1965). Table 4.2 gives transfer factor values, notes on some of the samples or test conditions and references. These factors have a median value of 0.0065 and a range from 0.00008 to 0.704.

It is notable that Kirchmann's transfer factors, based on field and greenhouse experiments, fall into the upper part of the range of transfer factors extracted from the literature and that pea pods had the highest value of any in his data set. McDowell-Boyer et al (1980) suggest that the loose sandy texture of the Belgian test soil may in part account for transfer factors above the North American average. In sandy soil relatively more of the Ra-226 present is available to be absorbed by plants while in more organic soil the radium tends to be bound and unavailable to plants.

4.2.4 Radium-226 from Soil to Root and Stem Vegetables

Root and stem vegetables showed low transfer factors for radium in all locations for which data were found. A summary of 24 data

Table 4.2

Radium-226 Transfer Factors: Vegetation (Above Ground) from Soil

Crop	TF*	Notes	Reference
cabbage, leaf	0.170	converted from dry weight basis as per McDowell-Boyer 1979	Kirchmann et al 1966
beans, pod	0.011		
beans, foliage	0.082		
peas, pod	0.704		
peas, foliage	0.136		
carrot, foliage	0.232		
beet, leaves	0.314		
potato, foliage	0.243		
broccoli	0.014	low calcium soil	U.S. NRC 1979
broccoli	0.005	high calcium soil	
cabbage, leaves	0.016		McDowell-Boyer et al 1979
swiss chard	0.008	control area	Tracy et al 1983
carrot, foliage	0.005		
beet, leaves	0.003		
cucumber	0.0004		
beans	0.0002		
swiss chard	0.010	contaminated area 1	
carrot, foliage	0.008		
beet, leaves	0.007		
cucumber	0.00008		
beans	0.0006		
carrot, foliage	0.004	contaminated area 2	
beet, leaves	0.002		
edible portions of mixed vegetables	0.0013	tomatoes, beans, lettuce, pumpkin, in equal parts	DeBortoli and Gaglione 1972
green vegetables	0.006	unrelated soil and vegetable samples in Germany	veg: Muth et al 1960 cited by Russell and Smith 1966; soil: Bundesminister 1961 cited by Klement 1965
	0.0007		
	0.016		
	0.0018		

* Transfer factor (TF) based on radium-226 content on a fresh weight plant/dry weight soil basis.

Table 4.3

Radium-226 Transfer Factors; Root Vegetables from Soil

Crop	TF*	Notes	Reference
carrots	0.066		Kirchmann et al 1966
beets	0.048		
potatoes	0.005		
potatoes	0.0007		U.S. NRC 1979
other below ground veg.	0.003		
turnips	0.011	low calcium in soil	Lindelcen 1977 cited by MacLaren 1979
	0.007	high calcium in soil	
carrots	0.001	control area	Tracy et al 1983
beets	0.002		
onions	0.0012		
potatoes	0.001		
carrots	0.0025	contaminated area 1	
beets	0.0009		
potatoes	0.0005		
carrots	0.005	contaminated area 2	
beets	0.003		
onions	0.004		
carrots	0.0007	contaminated area 3	
asparagus	0.0002		
potatoes	0.0006		DeBortoli and Gaglione 1972
root vegetables	0.004	unrelated soil and vegetable samples	veg: Muth et al 1960 cited by Russell and Smith 1966; soil: Bundesminister 1961 cited by Klement 1965
	0.0005		
	0.04		
	0.005		

* Transfer factor (TF) based on radium-226 content on a fresh weight plant/dry weight soil basis.

points obtained from 6 references is presented in Table 4.3. The median value of the data set is 0.00225. As evidenced from the data, the transfer factors for root vegetables are highly variable ranging between 0.0002 and 0.066.

Kirchmann's 1966 data again fall into the upper part of the transfer factor range. On the other hand, the German data from Muth et al (1960) and Bundesminister (1961), which use unrelated soil and vegetable concentrations, yield transfer factors that are not unlike those reported for Port Hope (Tracy et al 1983) where both contaminated and control garden plots and their produce were measured.

Corey et al (1983) found that most of the radium in root crops was incorporated into the consumable portion and not merely associated with surface skin and adhering soil. Hence, Tracy's values for washed vegetables (acid digested) are regarded as reliable. Corey's results, however, are all based on ashed weight (at 550°C) and are not comparable with the Port Hope data.

4.2.5 Radium-226 from Soil to Fruit

The data base reported in Table 4.4 for radium in fruit is quite limited and relies almost entirely on the observation of Tracy et al (1983) on Port Hope gardens. This may be the result of a general preoccupation during the past 25 years with transuranic elements released from nuclear tests. The very low observed transfer factors for radium may also be a factor. Further, most of the environmental assessments cited in this report contain no closely related data on fruit and soil because gardens or orchards did not exist at the mine sites.

Since fruit is formed mainly after the period of most active plant growth, it seems likely that the available mobile radium

Table 4.4

Radium-226 Transfer Factors: Fruit from Soil

Crop	TF*	Notes	Reference
raspberries	0.0014	contaminated area 4	Tracy et al 1983
raspberries	0.0068	contaminated area 5	
strawberries	0.0034	contaminated area 5	
apples	0.0009	control area	
apples	0.0007	contaminated area 2	
tomatoes	0.0005	contaminated area 1	
tomatoes	0.00012	contaminated area 2	
tomatoes	0.00008	contaminated area 4	
mixed fruit, edible portions	0.0005	equal parts of peaches, pears, apples, plums and grapes	DeBortoli and Gaglione 1972

* Transfer factor (TF) based on radium-226 content on a wet weight plant/dry weight soil basis.

supply is tied up in new leaves and stems and thus is not available for deposition in fruit. The low values for transfer factors reported in Table 4.4 clearly suggest that radium does not concentrate in fruit. As with the other vegetable crops, the observed transfer factors vary widely ranging from 0.00008 to 0.0068.

4.3 Uranium Transfer Factors

4.3.1 General

Uranium is ubiquitous in the air and widespread at trace levels in soils. Uptake from soil by vegetation depends on plant species (the highest levels being found in perennial plants) and the availability of the uranium in the soil (Garner 1972). It is believed that uranium enters plant roots mainly as the uranyl ion UO_2^{2+}(Hanson et al 1960 cited by Russell and Smith 1966).

The ratio of uranium in dry plant material to available uranium in soil may vary by a factor of 10 between individual plants and species but was found to be slightly above unity for grains, legumes, tubers, roots and hays (Garner 1972). Transfer factors in plants grown where soil and plant materials contain similar levels of uranium as a rule are in the 0.01 to 0.5 range, while for mosses they are close to 1.0 (Verkhovskaya et al 1972). The transfer factors reported for all vegetative types herein range between 0.00001 and 2.2.

4.3.2 Uranium from Soil to Grass, Hay and Herbage

This category of vegetation was chosen to include plants which might be used for cattle feed. A listing of the species which have been sampled is given in Table 4.5 with the exception of the identity of Hanson's 1976 "late spring growth", which is not

Table 4.5

Uranium Transfer Factors: Grass, Hay and Herbage from Soil

Plant Species	TF*	Notes	Reference
grasses 1	0.0050		Norcen 1980
2	0.0062		
3	0.0066		
4	0.0072		
5	0.018		
mixed grass	0.27	from field, near uranium mill	Whicker 1978 cited by MacLaren 1979
cheat grass	0.27	from field lysimeters	
mixed grass	0.002	calculated from mean values from 4 species grown on U tailings - samples washed prior to drying	Moffett and Tellier 1977
late spring growth	0.02 to 0.06	mainly green material	Hanson 1976 cited by MacLaren 1979
uranium indicator plants	0.01 to 1.0		
pasture grass	0.0025		U.S. NRC 1979
hay	0.0025		
grass	0.042		SENES 1982
mixed grasses	0.08	natural background area	Whicker and Ibrahim 1983
	0.03	reclamation area	
	1.1	exposed U tailings	
	2.2	U tailings pond edge	

* Transfer factor (TF) based on radium-226 content on a dry weight plant/dry weight soil basis.

known. Also since Hanson (1976 cited by MacLaren 1979) did not report median values, the upper and lower limits have been presented in the summary.

The transfer factors presented on the table are expressed as the ratio of uranium concentration in dry plant material to the concentration in dry soil. Not surprisingly, the factors cover a wide range from 0.0025 to 2.2 which equates to a 1000 fold difference between the minimum and maximum values.

4.3.3 Uranium from Soil to Vegetables (Above Ground)

Very few references were found which reported the concentration of uranium in above-ground vegetables and in related soil. Consequently, data for foliage of common vegetables were included in the summary given in Table 4.6. Some items, such as beet tops and pea pods are occasionally used as human food. Cucumbers, which contain about 96% water (Spector 1956 cited by McDowell-Boyer et al 1979) are sometimes classed as fruit, together with tomatoes, and sometimes as vegetables.

All transfer factors are far below unity and support the conclusion that uranium is not concentrated from soil by these plants. The median value of the 9 transfer factors is 0.001. The values range from 0.000012 to 0.007.

4.3.4 Uranium from Soil to Root and Stem Vegetables

Almost all of the transfer factors summarized in Table 4.7 for root and stem vegetables have been derived from the work of Tracy et al (1983). Several contaminated garden plots in Port Hope, Ontario were assigned numbers. These, plus a control plot, yielded soil and vegetable samples which were analysed for Ra-226, Pb-210 and uranium. Related soil and vegetable results

Table 4.6

Uranium Transfer Factors: Vegetables (Above Ground) from Soil

Crop	TF*	Notes	Reference
leafy produce			
- high	0.007		
- low	0.001		Bowen 1979 cited by Trabalka and Garten 1983
peas	0.003	all above ground plant parts	Schreckise and Cline 1980 cited by Trabalka and Garten 1983
vegetables	0.0025	edible above ground parts	U.S. NRC 1979
beet leaves	0.0045	control area	Tracy et al 1983
carrot leaves	0.00005	contaminated area 2	
beet leaves	0.00016	contaminated area 2	
cucumbers	0.000012	contaminated area 2	
carrot leaves	0.0005	contaminated area 6	

* Transfer factor (TF) based on radium-226 content on a wet weight plant/dry weight soil basis.

Table 4.7

Uranium Transfer Factors:
Root and Stem Vegetables from Soil

Crop	TF*	Notes	Reference
carrots	0.0002	contaminated area 1	Tracy et al 1983
carrots	0.0015	contaminated area 2	
carrots	0.0004	contaminated area 6	
onions	0.00045	contaminated area 2	
onions	0.00013	contaminated area 3	
potatoes	0.00089	control area	
potatoes	0.00002	contaminated area 1	
rhubarb	0.00008	contaminated area 3	
rhubarb	0.0005	contaminated area 4	
asparagus	0.00026	contaminated area 5	
below ground vegetables	0.0025		U.S. NRC 1979,

* Transfer factor (TF) based on radium-226 content on a fresh weight plant/dry weight soil basis.

permitted the calculation of credible transfer factors.

Generally, the factors given in Table 4.7 are only about one-tenth of the values given for the transfer factors for above-ground vegetables in Table 4.6. On the other hand, considering only the data reported by Tracy et al (1983), it is apparent that there is little difference between above-ground and root/stem vegetables. The values in Table 4.7 for root and stem vegetables range from 0.00002 to 0.0025. The median of the eleven values for root vegetables is 0.0004.

4.3.5 Uranium from Soil to Fruit

Only one source of data was found which gave concentrations of uranium in fruits and in related soil. Few of the sources reviewed discriminated sufficiently between food types, and if they listed concentrations or transfer factors, the focus was mainly on transuranic elements.

The Port Hope data given in Table 4.8, though somewhat scant, is useful for the purposes of this report. The data indicate that even where the soil concentration of uranium is high, fruit-soil transfer factors are far below unity. Interestingly, the transfer factors calculated for fruit taken from control areas were equivalent to or greater than the transfer factors determined on fruit samples taken from areas with elevated soil uranium levels.

4.4 Lead-210, Polonium-210 and Thorium-230 Transfer Factors

A very limited data base on lead-210, polonium-210 and thorium-230 levels in terrestrial vegetation was found in the literature. Table 4.9 presents a summary of transfer factor data for various grasses for each of the radionuclides.

Table 4.8

Uranium Transfer Factors: Fruit from Soil

Crop	TF*	Notes	Reference
raspberries	0.0018	contaminated area 4	Tracy et al 1983
raspberries	0.00026	contaminated area 5	
strawberries	0.00026	contaminated area 2	
apples	0.00039	control area	
apples	0.00029	contaminated area 3	
tomatoes	0.0004	control area	
tomatoes	0.00001	contaminated area 4	

* Transfer factor (TF) based on radium-226 content on a fresh weight plant/dry weight soil basis.

Lead-210 exists ubiquitously in the atmosphere as a long-lived daughter of radon-222. In addition, it is introduced into the atmosphere by numerous industrial activities. These sources contribute to an increase in naturally occurring soil levels. It is to be expected therefore, that the lead-210 content of plants will, in many cases, be attributable to aerosol-derived lead as well as soil-derived lead. Surprisingly, the transfer factor data reported on Table 4.9 vary over a narrow range from 0.028 to 0.42. The lead-210 data for grass compare favourably with data reported by McDowell-Boyer et al (1980) for forage, feed and hay of 0.02 to 0.3.

Like lead-210, polonium-210 is a radon decay product. It thus enters the biosphere and lithosphere largely through aerial deposition. The limited data presented in Table 4.9 show considerably more variability than that reported for lead-210. Polonium-210 transfer factors for grasses range from 0.003 to 1.10.

Thorium, like lead, occurs ubiquitously throughout the lithosphere. The mobility of thorium is generally quite low and its uptake by plants is generally considered to be negligible (Whicker and Schultz 1982). The transfer factors reported for grasses on Table 4.9 however indicate a wide range in thorium-230 uptake. Several of the measurements are for grasses grown on tailings, which may be a significant factor.

Table 4.9

Lead-210, Polonium-210 and Thorium-230 Transfer Factors: Grass from Soil

Plant Species	TF*		Notes	Reference
	A	B		
Lead-210				
grass, unidentified	0.28		mean values	Norcen 1980
blue grass	0.42		only reported	
mixed grasses, Wyoming	0.33		field at uranium mill	Whicker 1978
4 grass species, Ontario	0.04		grown on U tailings calculated from mean values for 4 species	Moffett and Tellier 1977
pasture grass	0.028			U.S. NRC 1979
stored feed (hays)	0.036			
Polonium-210				
grass (creeping red frescue, canaray grass, red top, chinese timothy)	0.003		TF calculated from mean value for all species	Moffett and Tellier 1977
mixed grasses	0.74	0.66	nat'l bkgd area	A) Whicker & Ibrahim
Wyoming	0.07	0.10	reclamation area	1983
	0.14	0.25	exposed tailings	B) Ibrahim et al 1982
	1.10	1.4	tailings pond edge	
Thorium-230				
mixed grasses	0.14		natural bkgd	Whicker and Ibrahim
	0.01		reclamation area	1983
	0.41		exposed tailings	
	10.0		tailings pond edge	
4 species of grass grown on U tailings Ontario	0.003		calculated from mean values for all 4 species	Moffatt and Tellier 1977 cited by MacLaren 1979
leafy vegetables	<0.0009-0.02		range	Bowen 1979 cited by Trabalka and Garten 1983

* Transfer factors (TF) based on radionuclide content on a dry weight plant/dry weight soil basis.

REFERENCES

Boone, F.W., and Ng, Y.C., 1981. "Terrestrial Pathways of Radionuclide Particulates." Health Physics, Vol. 41, pp. 735-747.

Canadian Standards Association (CSA), 1984. "Guidelines for Calculating Derived Release Limits for Radioactive Material in Airborne and Liquid Effluents for Normal Operation of Nuclear Facilities." Draft CSA Standard N288.1. January.

Corey, J.C., Boni, A.L. and Watts, J.R., 1983. "The Relative Importance of Intake and Surface Adherence in Determining the Radionuclide Contents of Subterraneous Crops." Health Physics, Vol. 44, pp. 19-28.

De Bortoli, M. and Gaglione, P., 1972. "Radium-226 in Environmental Materials and Foods." Health Physics, Vol. 22, pp. 43-48.

D'Souza, T.J. and Mistry, K.B., 1970. "Comparative Uptake of Thorium-230, Radium-226, Lead-210 and Polonium-210 by Plants." Radiation Botany, Vol. 10, pp. 293-295.

Garner, R.J., 1972. "Transfer of Radioactive Materials from the Terrestrial Environment to Animals and Man." CRC Press, Cleveland.

Hoffman, F.O., Bergstrom, U., Gyllander, C. and Wilkens, A.-B., 1984. "Comparison of Predictions from Internationally Recognized Assessment Models for the Transfer of Selected Radionuclides Through Terrestrial Food Chains." Nuclear Safety, Vol, 25, pp. 533-546.

Ibrahim, S.A., Flot, S.L. and Whicker, F.W., 1982. "Concentrations and Observed Behaviour of Radium-226 and Polonium-210 Around Uranium Mill Tailings." IAEA-SM-262/32. Reprint from "Management of Wastes from Uranium Mining and Milling."

Kirchmann, R., Boulenger, R. and Lafontaine, A., 1966. "Absorption Du Ra-226 Par Les Plantes Cultivees." Proc. First International Congress of Radiation Protection, Rome, Italy, Part 2, pp. 1045-1051.

Klement, A.W., 1965. "Natural Radionuclides in Foods and Food Source Materials." Chapter 6 in "Radioactive Fallout. Soils, Plants, Foods, Man." Edited by E.B. Fowler, Elsevier.

MacLaren, James F., Limited, 1979. "Radiological Monitoring of Uranium Fuel Processing Facilities." Prepared for the Department of Fisheries and Environment, Environment Canada, Ottawa.

McDowell-Boyer, L.M., Watson, A.P. and Travis, C.C., 1979. "Review and Recommendations of Dose Conversion Factors and Environmental Transport Parameters for Pb-210 and Ra-226 Final Report." Oak Ridge National Laboratory. NRC FIN. No. B0281.

McDowell-Boyer, L.M., Watson, A.P. and Travis, C.C., 1980. "A Review of Parameters Describing Terrestrial Food Chains Transport of Lead-210 and Radium-226." Nuclear Safety, Vol. 21, pp. 486-495.

Miller, C.W. and Hoffman, F.O., 1983. "An Examination of the Environmental Half-Time for Radionuclides Deposited on Vegetation." Health Physics, Vol. 45, pp. 731-744.

Mistry, K.B., Gopal-Ayengar, A.R. and Bharathan, K.C., 1965. "On the Radioactivity of Plants from the High Radiation Areas of the Kerala Coast and Adjoining Regions. II. Studies on the Uptake of Alpha and Gamma Emitters." Health Physics, Vol. 11, pp. 1459-1470.

Moffett, D., and Tellier, M., 1977. "Uptake of Radioisotopes by Vegetation Growing on Uranium Tailings." Canadian Journal of Soil Science,Vol. 57, pp. 417-424.

Ng, Y.C., Burton, C.A., Thompson, S.E., Tandy, R.K., Kretner, H.K. and Pratt, M.W., 1968. "Prediction of the Maximum Dosage to Man from the Fallout of Nuclear Devices - IV. Handbook for Estimating the Maximum Internal Dose from Radionuclides Released to the Biosphere." U.S. AEC Reports Lawrence Radiation Laboratory, University of California, Livermore.

Ng, Y.C., 1982. "A Review of Transfer Factors for Assessing the Dose from Radionuclides in Agricultural Products." Nuclear Safety, Vol. 23, pp. 57-71.

Norcen Energy Resources Limited, 1980. "Blizzard Uranium Project Environmental Impact Statement."

Peterson, H.T., 1983. "Terrestrial and Aquatic Food Chain Pathways." Chapter 5 in "Radiological Assessment. A Textbook on Environmental Dose Analysis." Edited by J.E. Till and H.R. Meyer, NUREG/CR-3332 ORNL-5968. NTIS, Springfield.

Russell, R.S. and Smith, K.A., 1966. "Naturally Occurring Radioactive Substances. The Uranium and Thorium Series." Chapter 17 in "Radioactivity and the Human Diet." Edited by R.S. Russell. Pergamon, Oxford.

SENES Consultants Limited, 1982. "Preoperational Monitoring Program for Blind River. Final Report." Prepared for Eldorado Resources Limited.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long-term Effects of Uranium Mill Tailings in Canada - Phase I." Prepared for the National Uranium Tailings Program, CANMET, EMR, Ottawa.

Sheppard, M.I., and Thibault, D.H., 1983. "Natural Uranium Concentration Factor Distributions for Ten Plant Species Indigenous to the Precambrian Shield." Atomic Energy of Canada Limited, Technical Record TR-220, Pinawa.

Trabalka, J.R. and Garten, C.T., 1983. "Behaviour of the Long-Lived Synthetic Elements and Their Natural Analogs in Food Chains." in "Advances in Radiation Biology. Vol. 10." Academic Press.

Tracy, B.L., Pranti, F.A. and Quinn, J.M., 1983. "Transfer of Ra-226, Pb-210 and Uranium from Soil and Garden Produce; Assessment of Risk." Health Physics, Vol. 44, pp. 469-477.

United States Nuclear Regulatory Commission (U.S. NRC) 1979. "Calculation of Annual Doses to Man from Routine Releases of Reactor Effluents for the Purpose of Evaluating Compliance with 10CFR Part 50, Appendix 1." Reg. Guide 1.109.

United States Nuclear Regulatory Commission (U.S. NRC), 1982. "Calculational Models for Estimating Radiation Doses to Man from Airborne Radioactive Materials Resulting from Uranium Milling Operations." Reg. Guide 3.51, March.

Verkhovskaya, I.N., Vavilov, P.O. and Maslov, V.J., 1972. "Distribution and Redistribution of Uranium, Radium and Thorium in Natural Biogeocenoses." (NTIS Translation RFP-Trans-162).

Whicker, F.W., 1978. "Biological Interactions and Reclamation of Uranium Mill Tailings." Proceedings of a Symposium on Uranium Mill Tailings Management, November 1978.

Whicker, F.W., and Shultz, B., 1982. "Radioecology: Nuclear Energy and the Environment." CRC Press, Boca Raton.

Whicker, F.W., and Ibrahim, S.A., 1983. "Radiological Investigations of Uranium Mill Tailings Systems." 4th Technical Progress Report for the Period October 1, 1982 through September 30, 1983. Report No. DOE/EV/10305-11.

Zach, R., 1978. "FOOD II: An Interactive Code for Calculating Concentrations of Radionuclides in Food Products." AECL Report 6305, Pinawa.

Zach, R., 1982(a). "LIMCAL: A Comprehensive Food Chain Model for Predicting Radiation Exposure to Man in Long-Term Nuclear Waste Management." AECL Report 6827, Pinawa.

Zach, R., 1982(b). "Parameter Values for the Long-Term Nuclear Waste Management Food Chain Model LIMCAL." AECL Report 6828, Pinawa.

DISCUSSION PAPER
ON
THE TRANSFER OF RADIUM FROM THE ATMOSPHERE TO PASTURE VEGETATION

By

F.O. Hoffman, R.H. Gardner, and B.G. Blaylock

for

SENES Consultants Limited

June 1985

Abstract

Three models of varying complexity for predicting the concentration of Ra-226 in pasture vegetation from continuous deposition of atmospheric particulates containing Ra-226 are described. Two of the models include processes describing the interception of wet and dry deposition by the surfaces of vegetation and the uptake roots of Ra-226 from the soil horizon while the third model accounts only for direct interception of particulate deposition. Uncertainties in the model parameters are included in the analysis through the use of frequency distributions. The models are evaluated probabilistically to assess the effects on the predicted vegetation concentrations for two scenarios: (1) after 50 years of continuous deposition; and (2) after equilibrium is established between the atmosphere and soil. The results show that the most complex model has the largest uncertainties but all three models predict comparable mean concentrations in the pasture vegetation for the scenarios examined. While data are not available to assess which of the three models is most suitable, it can be argued that the third model (direct interception only) should be used for assessment purposes when there is a continuous source of airborne particulate containing radium-226.

TABLE OF CONTENTS

	Page
1.0 INTRODUCTION	1
2.0 MODEL DESCRIPTION	2
3.0 METHODS	3
4.0 RESULTS	4
5.0 CONCLUSIONS	7
REFERENCES	

1.0 INTRODUCTION

In this report, three models are examined for predicting the concentration of Ra-226 in pasture vegetation as the result of the continuous atmospheric deposition of particles containing Ra-226. The effect of parameter uncertainty is included in this analysis through the use of appropriate distributions to describe the parameter values. The frequency distributions of parameter values are subsequently used to obtain, for each model, a frequency distribution of predicted vegetation concentrations, assuming a unit concentration of Ra-226 in air.

The transfer of Ra-226 from the atmosphere to pasture vegetation involves the processes of direct interception of wet and dry deposition by the surfaces of vegetation as well as the accumulation of Ra-226 in soil and its eventual uptake by roots. In Phase I of the Probabilistic Assessment Project (SENES 1984), the process of soil-to-plant transfer of Ra-226 was not included. The process of direct interception of radium depositing from the atmosphere was assumed to dominate the amount of Ra-226 in pasture vegetation.

In this present study, the process of root uptake of Ra-226 is examined to determine its potential importance to the total amount of Ra-226 contamination on and in pasture vegetation. The effect of root uptake is analysed under two scenarios:

(1) after fifty years of continuous deposition of Ra-226 from the atmosphere, and
(2) after equilibrium is established between Ra-226 in the atmosphere and soil.

2.0 MODEL DESCRIPTION

The three models considered in this report (and described in Table 1) are of decreasing levels of complexity, and increasing degrees of parameter aggregation. The amount of parameter aggregation is determined by the extent to which direct measurements can be made of the transfer of material from air to vegetation that implicitly include the combined effect of a number of discrete physical processes. Examples of aggregated parameters are air-to-vegetation concentration factors and air-to-vegetation mass transfer rates.

The structure of Model 1 is similar to models used by the U.S. Nuclear Regulatory Commission (U.S. NRC 1977), the U.S. Environmental Protection Agency (Moore et al. 1979), and the International Atomic Energy Agency (IAEA 1982). This model was developed with ten parameters. Model 2 has nine parameters and is similar to the first, with the exception of an empirical estimate of the air-to-vegetation mass transfer rate V_D and an aggregated parameter B_a that relates the steady-state concentration of Ra-226 in vegetation to the amount of Ra-226 per unit area in the root zone of soil. Model 3 has only one variable parameter. This parameter is an empirically derived air-to-vegetation concentration factor determined from time-averaged measurements of the concentrations of Be-7 and Pb-210 in air and in a variety of herbaceous species of vegetation in the vicinity of Oak Ridge, Tennessee (Bondietti et al. 1984).

In Model 2, the fractional interception of direct deposition, r, and the standing crop of vegetation, Y, are used in the second term of the equation to convert estimates of the initial vegetation concentration to estimates of the total areal deposition rate to the ground surface, given the relationship

Table 1

EQUATIONS FOR THE THREE MODELS USED TO CALCULATE THE VEGETATION CONCENTRATION OF RA-226 GIVEN A CONTINUOUS CONCENTRATION OF 1 pCi/m^3 IN AIR

Model 1:

$$C_v = X\ V_d \left\{ [r\ (1-\exp(-\lambda_v t_e)(Y\ \lambda_v)^{-1}] + [B_v(1-\exp\ (-\lambda_s t_v)\ (p\ \lambda_s)^{-1}] \right\}$$

Model 2:

$$C_v = X\ V_D \left\{ [(1-\exp(-\lambda_v t_e)(\lambda_v)^{-1}] + [Y\ B_a(1-\exp\ (-\lambda_s t_b)(r\ \lambda_s)^{-1}] \right\}$$

Model 3:

$$C_v = X\ CF_{a,v}$$

Notes:

(a) Model 1 is the traditional mathematical formulation of the air-to-vegetation transfer of radionuclides (IAEA 1982, U.S. NRC 1977, Moore et al. 1979).

(b) Model 2 uses an empirical mass-normalized deposition factor V_D to calculate the initial vegetation concentration due to direct atmospheric deposition and an aggregated parameter B_a that relates the concentration of Ra-226 in vegetation to the areal concentration in the root zone of soil.

(c) Model 3 is based on empirically determined equilibrium concentrations of submicron aerosols (Be-7 and Pb-210) in air and above-ground portions of several types of herbaceous vegetation.

(d) See Table 2 for nomenclature.

$$V_{d(total)} = V_D\ Y/r$$

Estimates of the parameter V_D have been derived from measurements of the combined process of wet and dry deposition of submicron aerosols containing Be-7 and Pb-210 (Bondietti et al. 1984) using a ratio of the initial rate of atmospheric deposition of these radionuclides on herbaceous vegetation (pCi/kg [dry wt].day) to their time-averaged concentrations in air (pCi/m^3).

In Model 3, it is assumed that the transfer of radium from the atmosphere will not be substantially different from that observed in humid climates for atmospheric aerosols containing Be-7 and Pb-210. It is also assumed that the amount of radium in exposed pasture vegetation will be dominated by the processes affecting the interception and retention of direct atmospheric deposition and that, by comparison, the amount taken up from soil will be negligible. The definition of each model parameter, the estimated distributions of parameter values, and the references consulted are given in Table 2.

3.0 METHODS

Uncertainties associated with model parameters are represented quantitatively by specifying appropriate frequency distributions. The distributions of parameter values given in Table 2 are intended to encompass the range of "true" parameter values. These parameter estimates are based on scientific judgement applied to values published in the literature and on data collected by the authors. The modelling results are statistical summaries of 400 Monte Carlo simulations, where one value per parameter per simulation is randomly and independently selected from the parameter frequency distribution prior to each model simulation. Thus, the 400 sets of parameter values produce 400 predictions of the concentration of Ra-226 in pasture vegetation.

Table 2

PARAMETER VALUES FOR THE PREDICTION OF THE TRANSPORT OF RA-226 FROM AIR-TO-VEGETATION

Parameter Description Symbol	Distribution	(Units) Values	References
Concentration of Ra-226 in pasture grass		(pCi/kg, dry wt.)	
C_v	calculated		
Concentration of Ra-226 in air		(pCi/m^3)	
X	constant	1.0	
Total (wet+dry) deposition velocity		(m/d)	Chambers (1985)
V_d	uniform	min. = 300, max. = 3000	
Air to vegetaton mass transfer		$(m^3/kg.d)$	Bondietti et al (1984)[b]
V_D	uniform	min. = 1000, max = 5000	
Fraction of atmospheric deposition intercepted by pasture vegetation		(unitless)	Hoffman & Baes (1979)
r	normal	min. = 0.10 max. = 0.40 mean = 0.47 s.d. = 0.30	
Average standing crop biomass of pasture vegetation (dry wt.)		kg/m^2	Judgmental estimate
Y	uniform	min. = 0.10, max. = 0.40	
Effective rate constant for the loss of Ra-226 from the surface of vegetation		$(days)^{-1}$	Hoffman et al. (1984)
λ_v	lognormal	geo. mean = 0.041, geo. s.d. = 1.6	
Average field exposure time for vegetation		(days)	Judgemental estimate
t_e	loguniform	min. = 15 max. = 200	

Table 2 (continued)
PARAMETER VALUES FOR THE PREDICTION OF THE TRANSPORT OF RA-226 FROM AIR-TO-VEGETATION

Parameter Description Symbol	Distribution	(Units) Values	References
Grass/soil concentration ratio		(unitless)	Bergstrom et al. (1984)[c]
B_v	loguniform	geo. mean = 0.025 geo. s.d. = 2.5	
Soil area to plant transfer factor		(m^2/kg)	Judgemental estimate[d]
B_a	loguniform	min. = 6×10^{-4} max. = 2.2×10^{-5}	
Effective areal density of the soil root zone (kg/m^2)			Hoffman et al (1982)
p	lognormal	geo. mean = 213, geo. s.d = 1.12	
Effective rate constant for the removal of Ra-226 out of the root zone of soil		$(days)^{-1}$	Bergstrom et al (1984)
λ_s	loguniform	min. = 2.2×10^{-6} max. = 3.3×10^{-4}	
Length of time soil has received airborne deposits		(days)	Scenario dependent
t_b	constant	1.83×10^4, infinity	
air-to-vegetation concentration factor		(m^3/kg.dry wt.)	Bondietti et al. (1984)[b]
$CF_{a,v}$	lognormal	geo. mean = 5.2×10^4, geo. s.d. = 1.6	

Notes

a) estimated for the combined processes of wet and dry deposition for a climate with approximately 100 cm of precipitation annually.

b) derived from estimates of transfer rates for Be-7 and Pb-210 to leaves and whole plants, assuming that the contribution from soil uptake is negligible.

c) concentration in the dry weight of vegetation divided by the concentration in the dry weight of soil.

d) an aggregated parameter that is not dependent on soil bulk density and sample depth.

The results from 400 trials on each model are presented as a geometric mean, a geometric standard deviation and an upper and lower bound estimate. In this paper, the upper bound estimate is obtained by multiplying the geometric mean by the square of the geometric standard deviation. The lower bound estimate is obtained by dividing the geometric mean by the square of the geometric standard deviation. The ranges given by the upper and lower bounds in Tables 3 and 4 and in Figure 1 and 2 are thus equivalent to an approximate 95% probability interval within a lognormal distribution. It is emphasized that the probability of the "true" value being outside this estimated range is considered to be much lower than 5%. By necessity, the distributions of parameter values (give in Table 2) are not intended to describe the actual variability of the parameter (an unknown quantity as has been described elsewhere (SENES 1984)); instead, the intent is for these distributions to encompass the likely extent of parameter uncertainty.

4.0 RESULTS

The results for the concentration of Ra-226 in vegetation per unit concentration in the atmosphere over a period of 50 years are presented in Table 3 and Figure 1. The results for the condition when the concentration of Ra-226 in air and soil have reached equilibrium are shown in Table 4 and Figure 2. These results show considerable overlap of the distributions of predicted values from each of the three models.

Model 1, the most complex model, has the largest uncertainty. This is probably due to independent parameter estimation. The independent estimation of parameters ignores potential interrelationships between parameters which may be correlated. Quantification of these potentially important relationships

Table 3

RESULTS OF THE CONCENTRATION IN VEGETATION PER UNIT CONCENTRATION IN AIR FOR AN EXPOSURE PERIOD OF 50 YEARS

	Ratio of Vegetation to Air Concentrations (C_v/X)				
Model No.	Geo. Mean (m^3/kg,dry)	Geo. S.D. (unitless)	Lower Bound[a] (m^3/kg,dry)	Upper Bound[b] (m^3/kg,dry)	Uncertainty Contribution[c]
1	50,800	2.80	6,500	397,000	soil-plant uptake (1%) V_d (31%) r (40%) Y (17%) λ_v (6%) t_e (4%)
2	54,000	1.82	16,000	177,000	soil-plant uptake (10%) V_D (52%) λ_v (21%) t_e (10%) r (6%)
3	52,000	1.60	20,000	130,000	$CF_{a,v}$ (100%)

Notes:

a) lower bound = (geo. mean)/(geo. s.d.)^2
b) upper bound = (geo. mean) x (geo. s.d.)^2
c) based on the square of the simple correlation coefficient calculated from a linear regression of the log-transfored parameter values and model predictions.

FIGURE 1

A COMPARISON OF MODEL PREDICTIONS IN $Bq \cdot kg^{-1}$ VEGETATION TO $Bq \cdot m^{-3}$ AIR ASSUMING SOIL IS CONTINUOUSLY EXPOSED TO RADIUM IN AIR FOR FIFTY YEARS

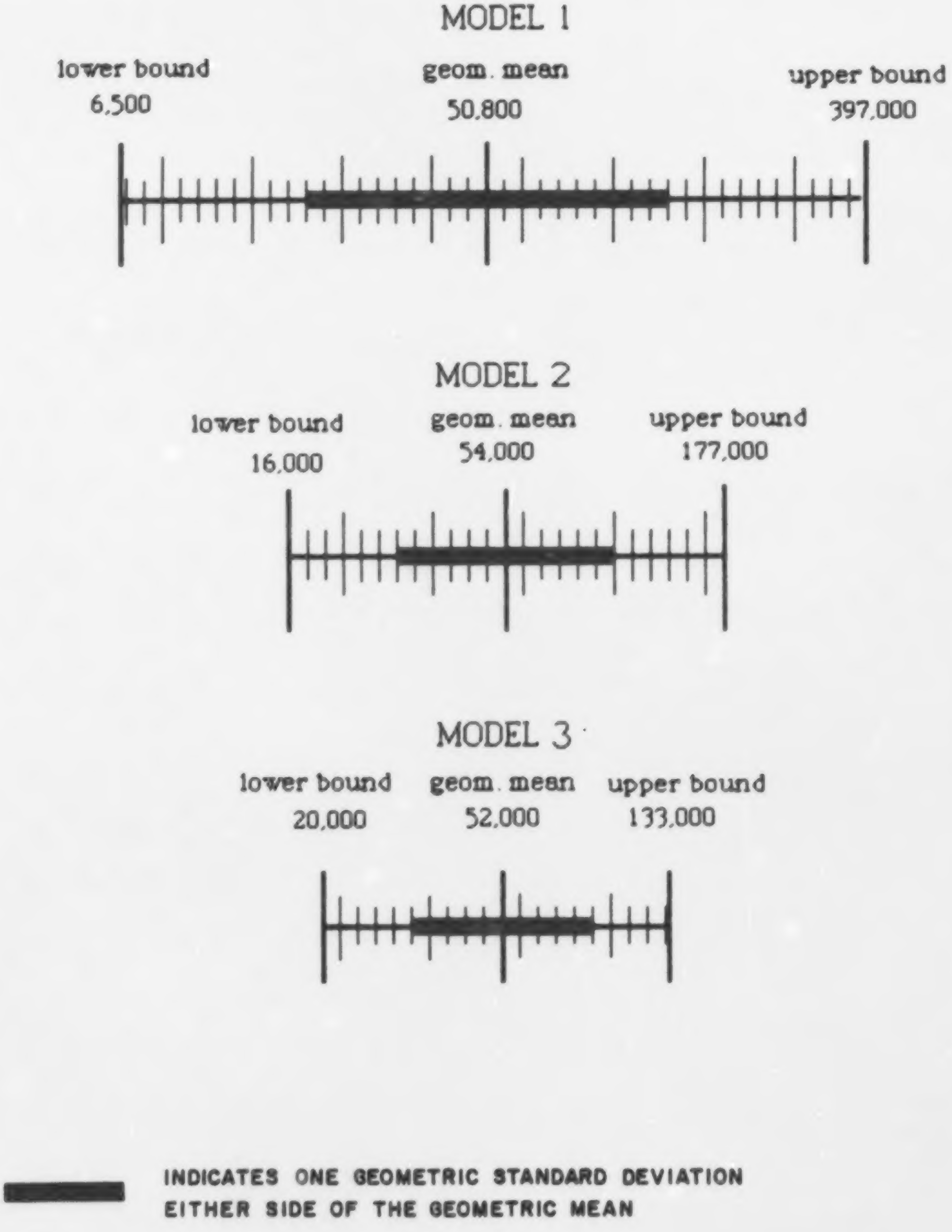


Table 4

PREDICTED VEGETATION CONCENTRATIONS FOR AN EXPOSURE PERIOD OF SUFFICIENT DURATION FOR EQUILIBRIUM TO OCCUR BETWEEN RA-226 IN AIR AND SOIL

Ratio of Vegetation to Air Concentrations (C_v/X)

Model No.	Geo. Mean (m^3/kg,dry)	Geo. S.D. (unitless)	Lower Bound[a] (m^3/kg,dry)	Upper Bound[b] (m^3/kg,dry)	Uncertainty Contribution[c]
1	64,900	2.80	8,300	510,000	soil-plant uptake (17%)
					V_d (34%)
					r (25%)
					Y (13%)
					λ_s (12%)
2	69,200	2.15	14,800	320,000	soil-plant uptake (30%)
					V_D (32%)
					r (16%)
					λ_s (14%)
					λ_v (9%)
3	52,000	1.60	20,000	130,000	$CF_{a,v}$ (100%)

Notes:

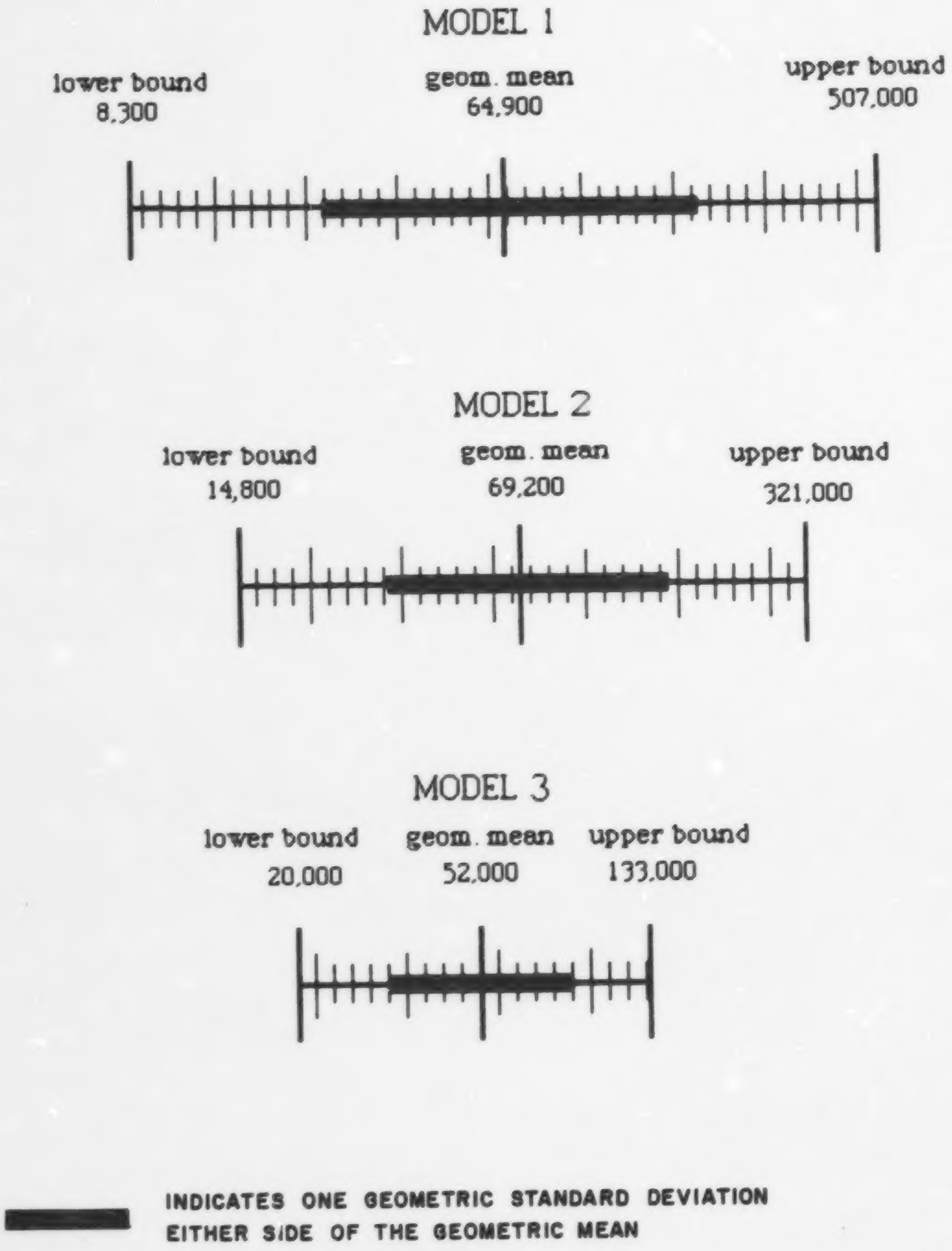
a) lower bound = (geo. mean) /(geo. s.d.)^2

b) upper bound = (geo. mean) x (geo. s.d.)^2

c) based on the square of the simple correlation coefficient calculated from a linear regression of the log-transformed parameter values and model predictions.

FIGURE 2

A COMPARISON OF MODEL PREDICTIONS IN Bq · kg^{-1} VEGATATION TO Bq · m^{-3} AIR ASSUMING EQUILIBRIUM BETWEEN THE CONCENTRATIONS OF RADIUM IN SOIL AND AIR



requires that the parameters be estimated using a system of simultaneous measurements.

Experience has shown that when parameter correlations are considered, the uncertainties in the model prediction will usually be less than when the model is constructed so that the model parameters are treated as statistically independent variables. The effect of parameter correlations is implicitly included within the aggregated parameters used in Models 2 and 3. Thus, the predicted values of these models are associated with tighter distributions than are obtained with Model 1 assuming that all of the parameters in Model 1 are statistically independent.

Although Model 3 does not include the process of radium uptake from soil, its predicted values are not substantially different from those of Model 2, and the geometric mean of the predicted values approximates that predicted by Model 1 for the case where the duration of soil exposure is 50 years (Figure 1). For the case where the duration of soil exposure is sufficiently long to produce an equilibrium between radium in air and soil, the uncertainty in Model 2 approaches the uncertainty exhibited by Model 1 (Figure 2), with the uncertainties of both models being noticeably larger than that for Model 3. The distribution of values predicted by Model 3 is encompassed by the range of values predicted by Models 1 and 2 in both the equilibrium case and the case where the soil is only exposed for 50 years.

For both Models 1 and 2, the contribution of the uptake of radium from soil to the total concentration of radium in vegetation increases when the duration of soil exposure is increased from 50 years out to the attainment of equilibrium between soil and air. The uncertainty contribution of the soil uptake pathways to the predicted vegetation concentration increases from less than 1% to

17% for Model 1 and from 10% to 30% for Model 2. Although the geometric mean values for both models, and the upper bound value for Model 1, increase by only about 8%, the upper bound value for Model 2 increases by more than 50%.

An analysis of the parameters contributing most to the uncertainty in the predictions of the models reveals that the most important parameters are those whose estimations are not based on specific observations for radium. These parameters are the atmospheric deposition terms, V_d and V_D, the interception fraction, r, the effective rate constant for the weathering of particulate matter from the surface of vegetation, λ_v, and the standing crop biomass, Y. These parameters are used to estimate the interception and retention of direct deposition of particulates by vegetation. However, vegetation interception and retention of aerosol deposition has not been shown to be appreciably different among a variety of metallic elements (Hoffman et al. 1982, Miller and Hoffman 1983, Bondietti et al. 1984), although some influence should be attributed to the predominant particle size of the depositing aerosol (Miller and Hoffman 1983).

It is suspected that if the deposition of particles containing radium were to be predominantly in the large particle size range (i.e. the super-micron range with particles larger than one micron in diameter), actual values of vegetation retention and interception might be appreciably less than the parameter values selected in this analysis. This is because of the possibility that vegetation interception and retention of super-micron particles may be substantially less efficient than for small (i.e. sub-micron) particles (Kirchner and Whicker 1984). However, the literature consulted for these parameters was not surveyed specifically for large particle sizes. If the vegetation efficiency for interception and retention of direct atmospheric deposition were less than assumed herein, the importance of the soil-to-plant pathway would increase.

5.0 CONCLUSIONS

The uncertainties associated with model predictions of the concentration of Ra-226 in pasture vegetation depend on the model used and the time period assumed for soil exposure to atmospheric deposition of radium. The simple aggregate model (Model 3) gave the least uncertain results (geometric standard deviation of 1.6), while the most complex model (Model 1) gave the most uncertain results (geometric standard deviation of 2.8). This difference may be attributed to a reduction in overall parameter uncertainty as parameters are aggregated in moving from the more complex to the simpler model structure.

Prior to the attainment of equilibrium between the concentrations of Ra-226 in air and soil, the uptake of Ra-226 from soil by vegetation is of negligible importance. At equilibrium, however, the importance of the uptake of Ra-226 from soil increases in both Models 1 and 2, although the processes of direct vegetation interception and retention of atmospheric deposition of Ra-226 still dominate the total concentration of Ra-226 in vegetation. Should a scenario be constructed where atmospheric deposition eventually ceased to be important, then Model 3 would, of course, be inadequate in this instance.

Conclusions about which of the three models is most correct must be postponed until comparisons can be made with independent sets of field data collected over prolonged periods of time. In the interim, conservative decisions could be based on the predictions of Model 1 since this model produces the highest upper bound estimate and the largest uncertainty. On the other hand, one could argue that Model 3 is the model that should be used for decision making, since: (a) it is the simplest of the three models; (b) it produces the smallest uncertainty; (c) the additional contribution of radium from the soil-to-plant uptake pathway is of minimal importance (assuming, of course, that atmospheric deposition continues over the duration of the scenario); (d) its geometric mean value is not appreciably different from the geometric mean values produced by the other two models; and (e) it implicitly includes the effect of correlations among the processes of wet and dry deposition and vegetation interception and retention of atmospheric aerosols.

In addition, all three models may be conservatively biased, given that the vegetation interception and retention of super-micron particles deposited downwind from uranium tailings sites may not be as efficient as is the interception and retention of micron to sub-micron particles. The data base reviewed to estimate the parameters describing vegetation interception and retention in Models 1 and 2 and the air-to-vegetation concentration factor in Model 3 was composed primarily of observations made for sub-micron to micron-sized particles. The extent and significance of this potential source of bias remains to be tested. Such testing could be accomplished when efforts are undertaken to compare model predictions against data obtained on the actual transfer of radium from air to vegetation downwind of uranium tailings.

REFERENCES

Bergstrom, U., Andersson, K., and Rojder, B., 1984. "Variability of Dose Predictions for Cesium-137 and Radium-226 Using the PRISM Method." Studsvik Energieteknik AB, Sweden, STUDSVIK/NW-84/656.

Bondietti, E.A., Hoffman, F.O. and Larsen, I.L., 1984. "Air-to-Vegetation Transfer Rates of Natural Submicron Aerosols." J. Environ. Radioactivity, Vol. 1, pp. 5-27.

Chambers, D.B., 1985. Personal communication. SENES Consultants Limited, Willowdale, Ontario.

Hoffman, F.O. and Baes, C.F. III (eds), 1979. "A Statistical Analysis of Selected Parameters for Predicting Food Chain Transport and Internal Dose of Radionuclides." ORNL/NUREG/TM-282.

Hoffman, F.O., Gardner, R.H., and Eckerman, K.F., 1982. "Variability in Dose Estimates Associated with the Food Chain Transport and Ingestion of Selected Radionuclides." NUREG/CR-2612, ORNL/TM-8099.

Hoffman, F.O., Miller, C.W., and Ng, Y.C., 1984. "Uncertainties in Radioecological Assessment Models." In "Seminar on the Environmental Transfer to Man to Radionuclides Released from Nuclear Installations." Commission of the European Communities, Oct. 17-21, 1983. Vol. 1, pp. 331-349, Brussels, Belgium.

International Atomic Energy Agency (IAEA), 1982. "Generic Models and Parameters for Assessing the Environmental Transfer of Radionuclides form Routine Releases: Exposures of Critical Groups." IAEA Safety Series No. 57.

Kirchner, T.B., and Whicker, F.W., 1984. Private communication on the status of the pathway analysis task for radionuclides released from the Nevada test site. Dept. of Radiobiology, Colorado State University, Fort Collins, Colo.

Miller, C.W. and Hoffman, F.O., 1983. "An Examination of the Environmental Half-time for Radionuclides Deposited on Vegetation." Health Phys., Vol. 45, pp. 731-744.

Moore, R.E., et al., 1979. "AIRDOS/EPA: A Computerized Methodology for Estimating Environmental Concentrations and Dose to Man from Airborne Releases of Radionuclides." ORNL-5532, Oak Ridge National Laboratory.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the the Long Term Effects of Uranium Mill Tailings in Canada - Phase I." Research Report Prepared for the National Uranium Tailings Program, CANMET, EMR, Canada.

U.S. Nuclear Regulatory Commission (NRC), 1977. "Calculation of Annual Doses to Man from Routine Releases of Reactor Effluents for the Purpose of Evaluating Compliance with 10 CFR Part 50." Regulatory Guide 1.109, Rev. 1.

Volume

3

NATIONAL
URANIUM
TAILINGS
PROGRAM

Energy, Mines and Resources Canada
Énergie, Mines et Ressources Canada

Canada

PROBABILISTIC MODEL DEVELOPMENT FOR THE ASSESSMENT OF THE LONG-TERM EFFECTS OF URANIUM MILL TAILINGS IN CANADA - PHASE II

VOLUME 3 - PROBABILISTIC ANALYSIS USER'S MANUAL AND DOCUMENTATION FOR THE URANIUM TAILINGS ASSESSMENT PROGRAM (UTAP)

A Research Report Prepared Under
Contract # OSQ84-00207 for

The National Uranium Tailings Program
Canada Centre for Mineral and Energy Technology
Energy, Mines and Resources Canada
555 Booth Street
Ottawa, Ontario
K1A 0G1

by

SENES Consultants Limited
499 McNicoll Avenue
Willowdale, Ontario
M2H 2C9

October 1985

ABSTRACT

This study was conducted to extend and refine the probabilistic assessment tool, developed in a previous study for assessing the long-term effects of uranium mill tailings disposal. The incorporation of time-dependent behaviour was a particularly important aspect of the study. Numerous technical issues relevant to performing or interpreting a probabilistic assessment were examined. A computer code was developed and used to perform probabilistic analyses on several uranium tailings management scenarios.

The study is described in three volumes. This volume, Volume 3, contains the user's manual for the probabilistic assessment code developed in the study and provides detailed documentation of the algorithms used in modelling the reference tailings site.

NOTICE

This report was prepared by SENES Consultants Limited (SENES) under contract to the National Uranium Tailings Program (NUTP). Neither the NUTP nor SENES makes any warranty, expressed or implied, or assumes any liability with respect to any damage or loss incurred as a result of the use made of the information contained in this publication.

TABLE OF CONTENTS

Page

Abstract i
Notice i

1.0 INTRODUCTION 1-1
  1.1 Background 1-1
  1.2 The Concept of Probabilistic Analysis 1-1
  1.3 Scope of This Report 1-3

2.0 CONCEPTUAL OVERVIEW OF THE SYSTEMS APPROACH 2-1
  2.1 Conceptual Considerations 2-1
  2.2 Internal Structure of the Random Simulation (RANSIM) Module 2-2
  2.3 Internal Structure of the Parameter Allocation and Control (PAC) Module 2-3

3.0 INPUTS, OUTPUTS AND INTERFACE WITH RANSIM 3-1

4.0 A SIMPLE GENERIC EXAMPLE 4-1
  4.1 Illustrative Example 4-1
  4.2 Details of the PAC Module 4-2
  4.3 Structure of the Simulation Module 4-4
  4.4 Coding of PAC and EXAM 4-4
  4.5 Compile and Link 4-5
  4.6 Input Data File and Parameter Name File 4-5
  4.7 "Running" the Program and Sample Results 4-6

5.0 THE URANIUM TAILINGS ASSESSMENT PROGRAM 5-1

6.0 QUALITY ASSURANCE 6-1

7.0 REFERENCES 7-1

Table of Contents (continued)

APPENDIX A - The Calculation of the Water Source Term (SUBROUTINE SOURCE)

APPENDIX B - The Calculation of Air Source Terms

APPENDIX C - Surface Water Dispersion Component Models

APPENDIX D - Air Dispersion Component Model

APPENDIX E - The Calculation of the Concentration of Radionuclide i in Environmental Media

APPENDIX F - The Calculation of the Annual Dose Received by the Critical Receptor

APPENDIX G - Groundwater Transport Model

LIST OF TABLES

Table No.		Follows Page
3.1	RANSIM First Input File Characteristics	3-1
3.2	RANSIM Distribution Descriptors	3-1
3.3a	Example of RANSIM Input Distributions	3-1
3.3b	Example Input Data Set	3-1
3.4	RANSIM Second Input File Characteristics	3-2
3.5	Example Interactive Session with RANSIM	3-2
4.1	Parameter Definitions for General Example	4-2
4.2	PAC, ENDOUT, TIMOUT and EXAM Subroutines	4-4
4.3	STUFF.INC File	4-5
4.4	Input Parameters	4-5
4.5	Output Parameter Name File	4-5
4.6	Printout of Three Trials for the EXAMS Model	4-6
4.7	Summary Statistics from Ten Trials	4-6
5.1	Input Parameters That Will Fill COMMON/INPT/	5-1
5.2	Parameter Name File for Parameters That Will Fill COMMON/RESULT/	5-1
5.3	Source Code for PAC, TIMOUT and ENDOUT	5-1
5.4	$INCLUDE File That Contains COMMON Blocks Used in PAC and Other Subroutines	5-1
5.5	Variable Dictionary for PAC, ENDOUT and TIMOUT	5-1

LIST OF FIGURES

Figure No.		Follows Page
1.1	Concept of a Deterministic Analysis	1-2
1.2	Concept of a Probabilistic Analysis	1-2
2.1	Conceptual Flow Diagram of Systems Approach	2-1
2.2	Conceptual Flow Diagram of RANSIM	2-2
2.3	Conceptual Flow Diagram of PAC	2-3
3.1	Input Information for Arbitrary Distributions	3-1
3.2	Example of an Arbitrary Distribution	3-1
4.1	Subroutine PAC for Simple Example	4-2
4.2	Subroutine EXAM	4-4
5.1	Conceptual Structure of the Uranium Tailings Assessment Program	5-1
5.2	Procedure Used to Change Parameters Written to Output File	5-2

1.0 INTRODUCTION

1.1 Background

In 1982 Energy, Mines and Resources Canada established the National Uranium Tailings Program (NUTP) which is intended to pull together previous work on Canadian uranium mill tailings, to fund additional work where significant information gaps are found, and to develop the capability to predict the long-term effects of tailings. The latter aspect requires the use of mathematical models to simulate the behaviour of the real physical system(s). In this regard, SENES Consultants Limited carried out a Phase I study for the NUTP to demonstrate how probabilistic techniques could be used to take account of the natural variability and uncertainty in model parameters in simulation analysis (SENES 1984).

Subsequently, SENES undertook a Phase II study to extend and refine the probabilistic tools developed in Phase I and to investigate various issues of technical interest involved in carrying out and interpreting a probabilistic analysis. The findings of the Phase II study are presented in a three-volume series of reports. This report, Volume 3 of the Phase II report, documents the structure and use of the system model used to carry out the probabilistic analysis and presents documentation for the simulation models used in the analysis.

1.2 The Concept of Probabilistic Analysis

Evaluating the potential long-term effects of uranium mill tailings requires mathematical models which simulate the environmental transport and exposure pathways of the various uranium tailings constituents. By their very nature, models are only partial representations of real systems and contain a number of parameters which are either imperfectly known or inherently variable by nature. Uncertainties also arise when these models are used to predict effects over conditions and periods of time different from those for which the models and model parameters were developed. Consequently, the results of all model calculations are uncertain to some degree.

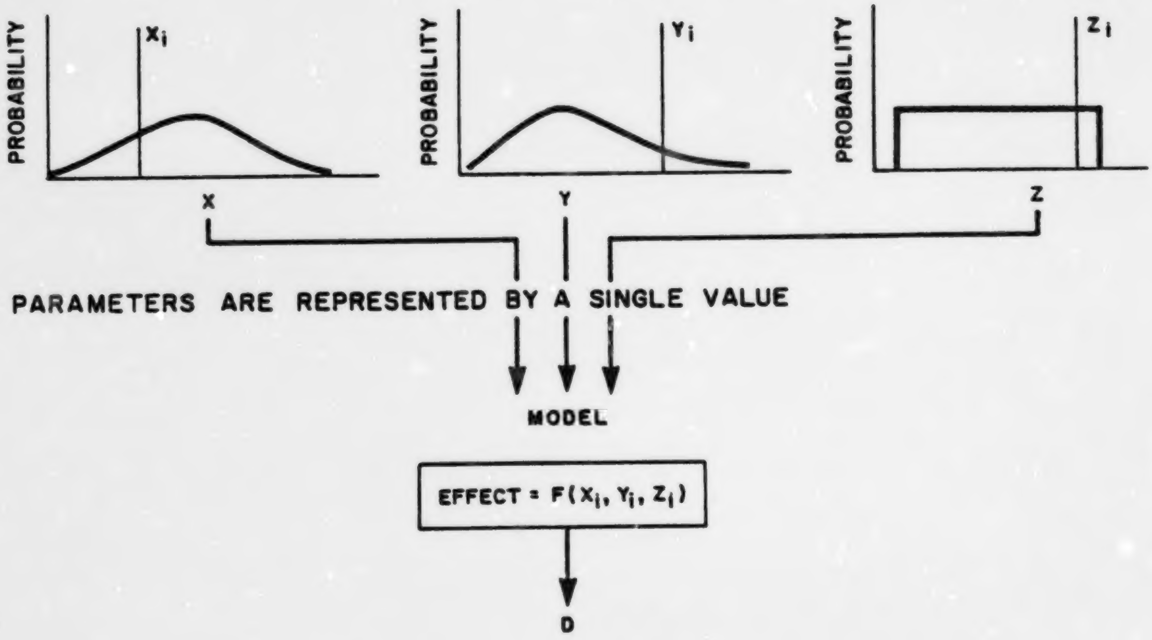
The traditional, deterministic approach to evaluating environmental effects is illustrated in Figure 1.1. In this approach, a single value is selected for each of the model parameters and a single value (the effect) is calculated in some suitable unit such as concentration, dose or risk. As suggested on the figure, conservative values of the parameters, typically near the upper or lower extreme of the expected range of the probability distributions for the parameters, are often selected. The result of such an approach is generally a very conservative estimate of the "effect".

A probabilistic analysis uses the same mathematical models as the deterministic analysis; however, the approach attempts to account explicitly for uncertainty in many of the factors involved. This is achieved by specifying a probability distribution rather than a single value for parameters input to the model. The model is then run many times with new values selected for the input parameters on each run. Thus, the result of a probabilistic analysis is a distribution of "effects". The parameter values used in each run (also referred to as a trial) are randomly selected from the appropriate distributions using a computer random sampling technique. With this approach, as many individual estimates of the "effect" are obtained as desired by increasing or decreasing the number of runs. The basic output from such an analysis is a frequency distribution of possible effects. The concepts of a probabilistic analysis are shown in Figure 1.2.

FIGURE 1·1

CONCEPT OF A DETERMINISTIC ANALYSIS

- A RANGE OF INPUT VALUES



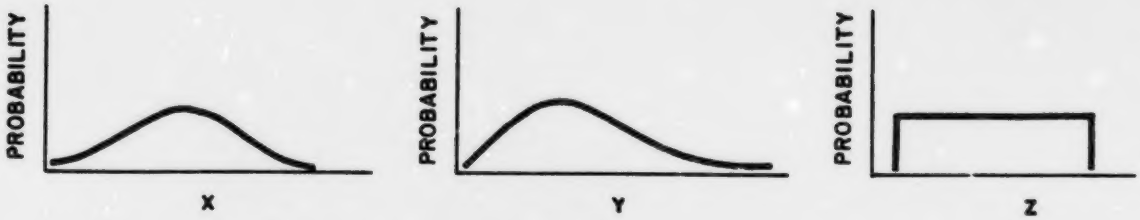
- PARAMETERS ARE REPRESENTED BY A SINGLE VALUE

- A SINGLE VALUE IS PRODUCED AS MODEL OUTPUT

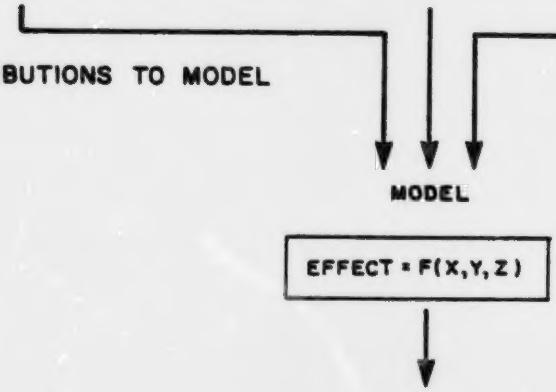
FIGURE 1·2

CONCEPT OF A PROBABILISTIC ANALYSIS

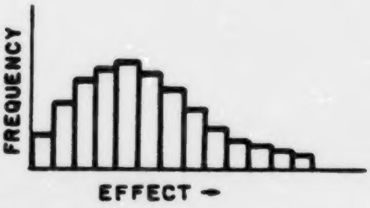
- DEVELOP PARAMETER DISTRIBUTIONS



- INPUT DISTRIBUTIONS TO MODEL



- PRODUCE A DISTRIBUTION OF MODEL PREDICTIONS



1.3 Scope of This Report

The objectives of this document are:

- to introduce the conceptual aspects and mechanical details of a probabilistic analysis,
- to permit a user to evaluate a mathematical simulation model in a probabilistic manner, and
- to act as a guide to the use of the specific model UTAP (Uranium Tailings Assessment Program).

The document can be read on several levels. It can serve as a conceptual guide for the less experienced computer user to explain how a model is evaluated in the probabilistic mode. Understanding the conceptual algorithms involved in such a process is an excellent way of familiarizing oneself with the logic involved in the analysis. The report can also serve as a "how to" guide to the more experienced computer user, furnishing the basic information necessary to integrate a simulation model into the probabilistic analysis process. Finally, the report provides support documentation and a user's manual for UTAP.

Section 2 introduces the reader to the logical systems approach utilized in developing the simulation models and provides a general description of the control modules. Section 3 describes the inputs, outputs, and techniques used to interface the various subroutines. Section 4 presents a simple example. Section 5 discusses UTAP and Section 6 provides a brief discussion of quality assurance activities in Phase II.

The mathematical formulations of the various models used in UTAP are presented in the appendices to this volume. The names of the parameters used in the UTAP computer code and their input distributions are also documented in the appendices.

2.0 CONCEPTUAL OVERVIEW OF THE SYSTEMS APPROACH

2.1 Conceptual Considerations

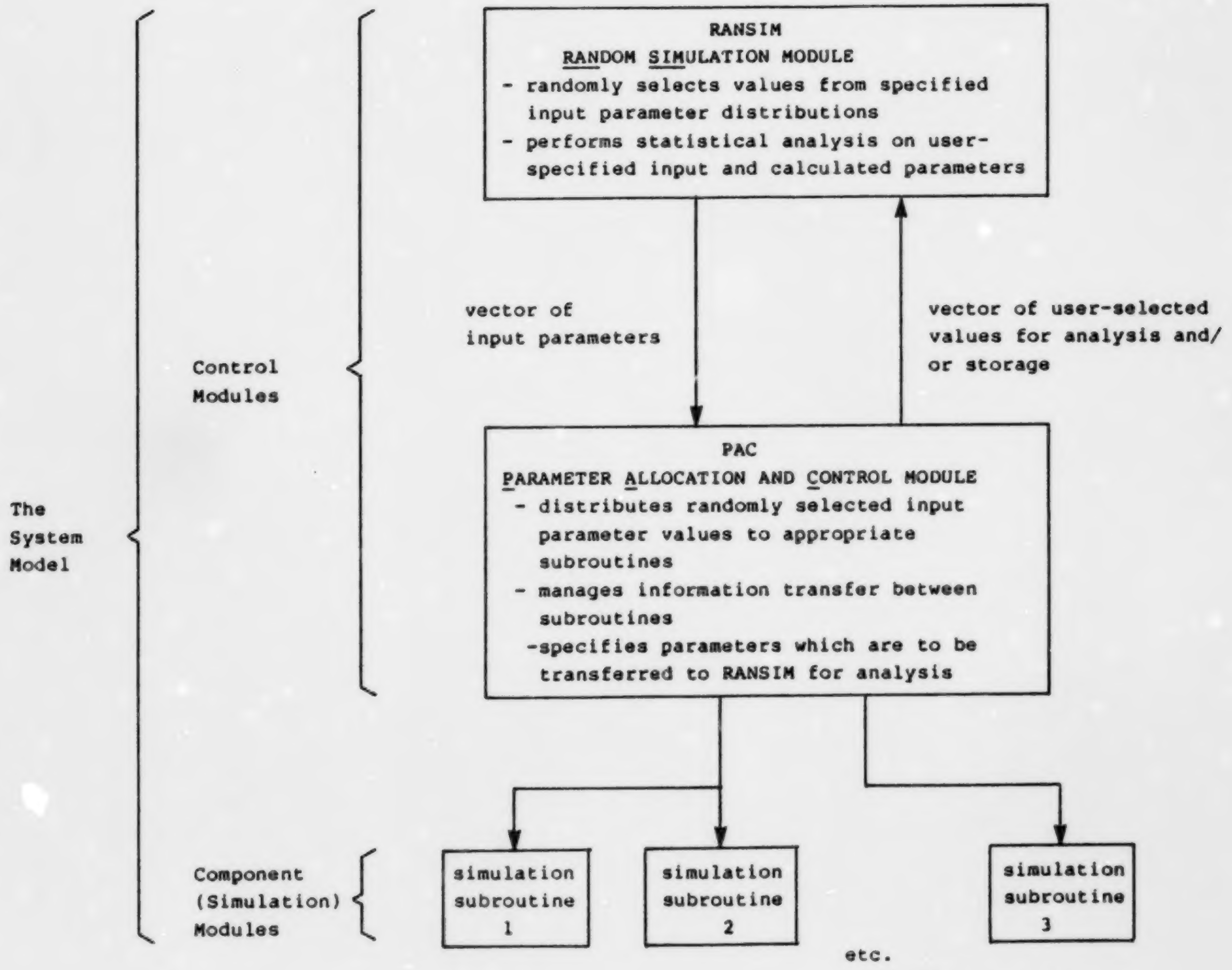
The probabilistic analysis process is inherently large and complex, hence a logical, step-by-step approach is required to make it tractable. The process is exceptionally well-suited to the application of a "top-down" computer programming approach. The top-down approach first involves identification of the overall system requirements, followed by the development of comprehensive, conceptual structures consistent with these requirements, followed in turn by further development of conceptual substructures. The overlying conceptual structures always dictate the nature of lower-level substructures. Once the conceptual algorithms are developed in this top-down manner, implementation is often a fairly mechanical programming exercise.

By way of example, consider the overall requirements for performing a probabilistic analysis. These include the ability to draw random samples from specified probability distributions, the ability to route these values to appropriate, user-specified simulation models, and the ability to store and/or analyze the large quantity of results generated from the many trials in a probabilistic analysis. The input values must be sampled from the appropriate distributions and, if several simulation models are being used, they must be evaluated in the correct order and pass any intermediate, calculated values between themselves correctly.

Given these overall requirements, the structure shown in Figure 2.1 was developed. As indicated on the figure, parameter allocation and control is performed by the PAC module. PAC acts as an intermediary between the sampling and statistics module (RANSIM) and the user-specified simulation subroutines. The

FIGURE 2·1

CONCEPTUAL FLOW DIAGRAM
OF SYSTEMS APPROACH


The System Model
Control Modules
RANSIM
RANDOM SIMULATION MODULE
- randomly selects values from specified input parameter distributions
- performs statistical analysis on user-specified input and calculated parameters
vector of input parameters
vector of user-selected values for analysis and/or storage
PAC
PARAMETER ALLOCATION AND CONTROL MODULE
- distributes randomly selected input parameter values to appropriate subroutines
- manages information transfer between subroutines
-specifies parameters which are to be transferred to RANSIM for analysis
Component (Simulation) Modules
simulation subroutine 1
simulation subroutine 2
simulation subroutine 3
etc.

internal structure of these modules is discussed below.

2.2 Internal Structure of the Random Simulation (RANSIM) Module

The random simulation (RANSIM) program is a variation of the OPANAL program originally developed by Dr. P.M. Reilly (Reilly and Johri 1969). The program has the capability to draw random values from a variety of probability distributions, route these values to a user-specified simulation model, and perform simple statistical analysis of selected results. RANSIM also maintains a file of output results which can be exported for post-processing on user provided graphical and statistical programs.

The internal structure of RANSIM is detailed in Figure 2.2. RANSIM is designed to convert numeric descriptors of many types of distributions into a series of randomly-sampled values available for transfer to simulation models. The distributions currently available are as follows:

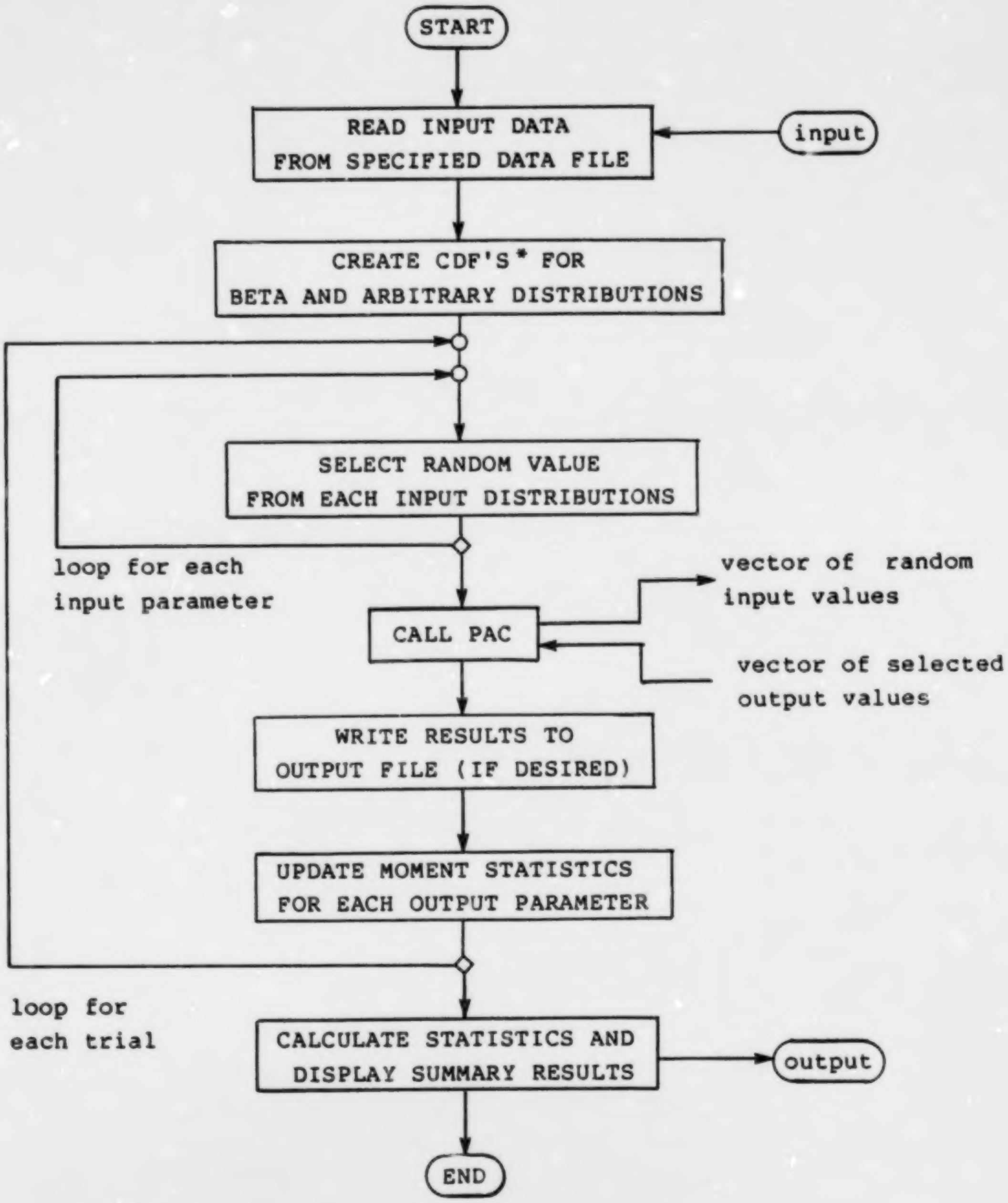
constant value
normal
lognormal
uniform
log-uniform
triangular
beta
correlated
arbitrary

The beta and arbitrary distribution capabilities allow almost an infinite range of distributions to be specified.

RANSIM has the capacity to calculate rapidly a variety of summary statistics on either user-specified input parameters or calculated parameters. Summary statistics available include:

FIGURE 2·2

CONCEPTUAL FLOW DIAGRAM OF RANSIM


START
READ INPUT DATA FROM SPECIFIED DATA FILE
input
CREATE CDF'S* FOR BETA AND ARBITRARY DISTRIBUTIONS
SELECT RANDOM VALUE FROM EACH INPUT DISTRIBUTIONS
loop for each input parameter
CALL PAC
vector of random input values
vector of selected output values
WRITE RESULTS TO OUTPUT FILE (IF DESIRED)
UPDATE MOMENT STATISTICS FOR EACH OUTPUT PARAMETER
loop for each trial
CALCULATE STATISTICS AND DISPLAY SUMMARY RESULTS
output
END

* CUMULATIVE DISTRIBUTION FUNCTION

arithmetic mean
arithmetic standard deviation (SD)
geometric mean (GM)
geometric standard deviation (GSD)
skewness
kurtosis
maximum value (max)
minimum value (min)

These are all calculated using recursive ("on-the-fly") methods and thus can be determined regardless of whether or not all values from all trials are to be saved. The summary statistics may be most beneficial when the amount of information generated is large and, if stored, would present computer storage problems.

The transfer of data between RANSIM and the simulation subroutines is achieved by the PAC module.

2.3 Internal Structure of the Parameter Allocation and Control (PAC) Module

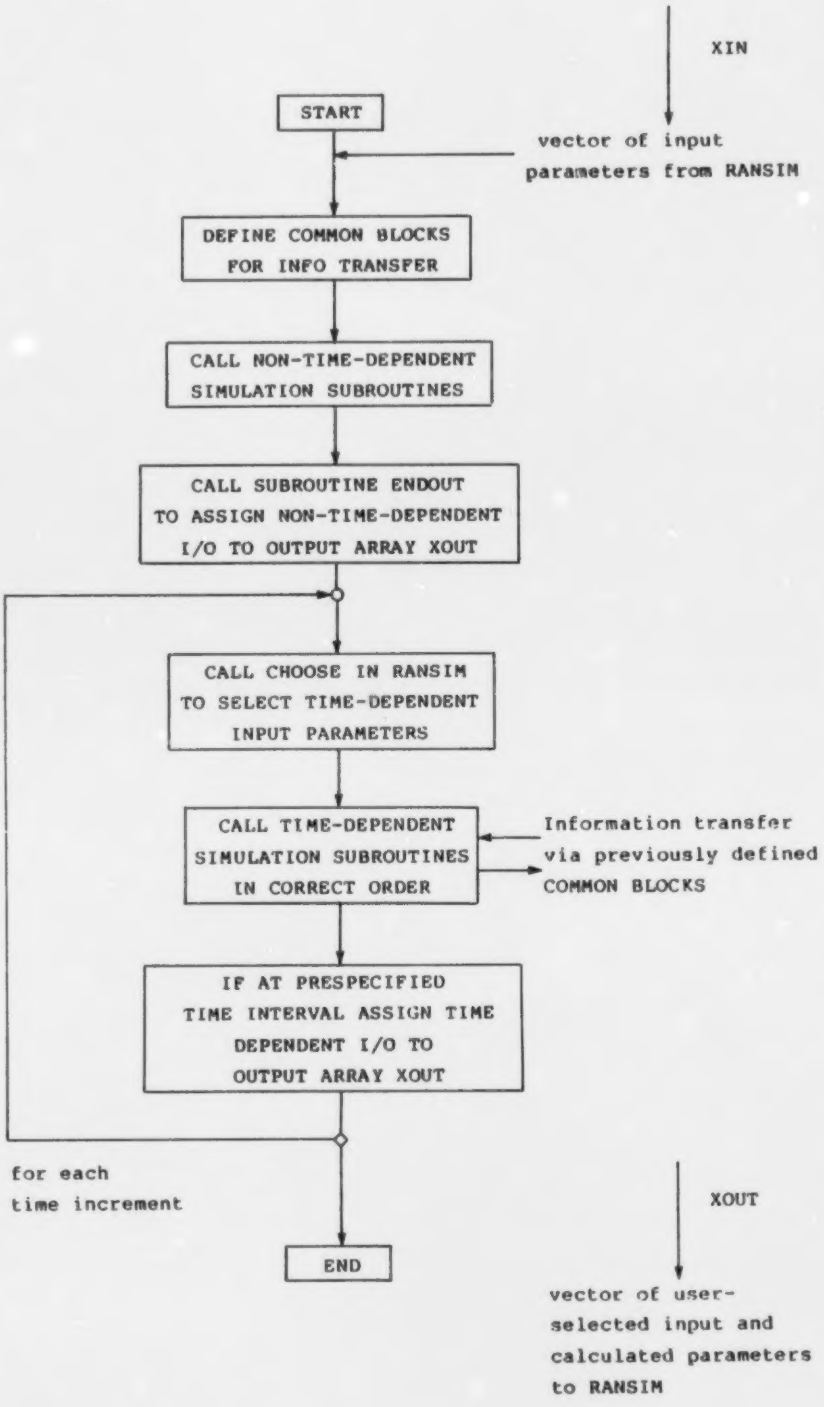
For ease of implementation of a probabilistic analysis, the most important conceptual substructure is the PAC module. Its structure at a broad, conceptual level is defined in Figure 2.3.

As indicated on Figure 2.3, information transfer between RANSIM, PAC, and the various simulation subroutines is achieved via the use of "labelled COMMON" statements. Such a statement causes the computer system to create a set of storage locations (called the COMMON blocks) that are accessible to all subroutines containing that particular labelled COMMON statement. For example, the labelled COMMON statement in one subroutine:

```
COMMON/INPT/XIN(5)
```

FIGURE 2·3

CONCEPTUAL FLOW DIAGRAM OF PAC


XIN
START
vector of input parameters from RANSIM
DEFINE COMMON BLOCKS FOR INFO TRANSFER
CALL NON-TIME-DEPENDENT SIMULATION SUBROUTINES
CALL SUBROUTINE ENDOUT TO ASSIGN NON-TIME-DEPENDENT I/O TO OUTPUT ARRAY XOUT
CALL CHOOSE IN RANSIM TO SELECT TIME-DEPENDENT INPUT PARAMETERS
CALL TIME-DEPENDENT SIMULATION SUBROUTINES IN CORRECT ORDER
Information transfer via previously defined COMMON BLOCKS
IF AT PRESPECIFIED TIME INTERVAL ASSIGN TIME DEPENDENT I/O TO OUTPUT ARRAY XOUT
for each time increment
XOUT
END
vector of user-selected input and calculated parameters to RANSIM

could be referenced in another subroutine by:

```
COMMON/INPT/A,B,C,D,E
```

In other words, XIN(1) would correspond to A, XIN(2) to B and so on.

The concept of time is managed by PAC, which permits the simulations to be carried out over varying time periods. Inputs to time-dependent simulation models can be divided into three major categories.

1) Physical constants (parameters) that are always fixed at the same value such as radioactive decay constants and Avogadro's number.

2) Uncertain, non-time-varying parameters that remain fixed for an entire simulation time period but vary stochastically from trial to trial such as soil bulk density and the air-to-vegetation transfer factors.

3) Uncertain, time-variant parameters that vary from one time step to the next such as annual precipitation and mean annual wind speed.

The first two categories of parameters are sampled only once for each trial. Uncertain time-variant parameters, on the other hand can be treated in several ways and require special consideration. Different ways of treating parameters which vary from year to year have been discussed in Section 2.6 of Volume 1 and in the discussion paper by Beals et al. (Appendix D) of Volume 2.

At the present stage of model develoment, time-variant parameters are sampled once each trial at the same time the uncertain non-time-variant parameters are sampled. In this case the variance of the sampling distribution of mean annual values of time-variant parameters is adjusted by a factor of 1/N where the simulation is carried out over N years.

Non-time-dependent simulation models only have input parameters from categories one and two and are evaluated with an effective time loop of one time step.

3.0 INPUTS, OUTPUTS AND INTERFACE WITH RANSIM

RANSIM is designed to accept, store, and transfer information, and interface with user-defined programs in a specific manner. The inputs to, and interface with, RANSIM are discussed here in generalities but in sufficient detail for a user to understand the input requirements of the worked example presented in Section 4 and the UTAP description provided in Section 5. Some computer programming experience on the reader's part is assumed. Since RANSIM is an interactive program an example is also given of an interactive session.

The FORTRAN code for the UTAP system model currently resides in an IBM PC AT utilizing the IBM FORTRAN version 2.0 compiler. Hence the filenames and certain code characteristics are specific to this hardware and software. However, only minimal changes would be required to make the code compatable with other environments.

Most information supplied to RANSIM is contained in two input data files. The name of the first file (including its three character extention) must be 12 characters long (e.g. EXAMPLE1.DAT). This file contains input parameter descriptions and other necessary information and is described in detail in Tables 3.1 and 3.2 and illustrated in Figures 3.1 and 3.2 for an arbitrary distribution. Examples of how each of the nine available distributions are actually implemented are provided in Table 3.3. Table 3.3(a) and Figure 3.2 describe distributions for each of nine variables (arbitrarily specified as VAR1 to VAR9). Each variable in this example corresponds to a different type of distribution. Table 3.3(b) represents the actual input data file corresponding to the descriptions of Table 3.3(a) and Figure 3.2. Table 3.3(b) also indicates:

Table 3.1

RANSIM First Input File Characteristics

	Column No.	Format	Parameter Description
Line 1	1-40	A40	Data set Title
Line 2	1-4	I4	No. of beta distributions (NIN)
	5-8	I4	No. of input parameters (NPARM)
	9-12	I4	No. of cells in beta distribution(s) (NCELLS)
	13-16	I4	No. of correlated parameters (NCORR)
	17-20	I4	No. of arbitrary distributions (NARB)
Line 3			
	1-10	I10	Random Number Seed (NSTART)
	11-14	I4	Number of iterations

Line 4 to (3 + the number of input parameters)

First specify all parameters that are constant, normal, lognormal, uniform, loguniform, triangular or beta distributed: (see Table 3.2 for details)

1-6	A6	Parameter name
9-10	I2	Distribution number (see Table 3.2)
11-54	4E11.6	Parameter distribution descriptors (see Table 3.2)

Next, specify all correlated parameters:

1-6	A6	Parameter name
9-10	I2	Distribution Number (8)
11-76	6E11.6	Parameter Distribution Descriptors (see Table 3.2)

Finally, specify all arbitrary parameters

1-6	A6	Parameter name
9-10	I2	Distribution number (9)
11-13	I3	Number of discrete cells in arbitrary distribution (NACELL) (max 25)
14-35;	2E11.6	Minimum value, maximum value
36;	NACELL*F6.2	Ordinates of probability density Function for each cell (sequentially from smallest to largest)

Table 3.2

RANSIM Distribution Descriptors

Distribution	Distribution Number	Distribution Descriptors	Column
constant	1	constant	11-21
normal	2	mean	11-21
		standard dev.	22-32
		min.	33-43
		max.	44-54
lognormal	3	geometric mean	11-21
		geo. stand. dev.	22-32
		min.	33-43
		max	44-54
uniform	4	min	11-21
		max	22-32
log-uniform	5	min.	11-21
		max	22-32
triangular	6	min	11-21
		peak	22-32
		max	33-43
beta	7	min	11-21
		max	22-32
		a	33-43
		b	44-54
correlated	8	mean of dependent variable	11-21
		S.D of dependent variable	22-32
		number of independent variable	33-43
		correlation coefficient	44-54
		mean of independent variable	55-65
		S.D of independent variable	66-76
arbitrary	9	number of cells	11-13
		min	14-24
		max	25-35
		ordinate values (see Figure 3.1)	36-(NACELL *6)

Table 3.3(a)

Example of RANSIM Input Distributions

Variable Name	Type	Description
Var1	constant	10.0
Var2	normal	mean = 5.0; SD = 1.0 min = 2.0; max = 8.0
Var3	log-uniform	min = 1.0 max = 1000
Var4	triangular	min = 0.0 mode = 5.0 max = 10.0
Var5	lognormal	GM = 50; GSD = 2.0 min = 6.0; max = 400.0
Var6	uniform	min = 1.0 max = 10.0
Var7	beta	min = 5; max = 25 a = 2.0; b = 2.5
Var8	correlated with Var2	mean = 1.0; SD = 0.2 correlation coeff = 0.8
Var9	arbitrary (see Figure 3.2)	

Table 3.3(b)

Example Input Data Set

```
PROBABILISTIC ANALYSIS EXAMPLE DATASET
   1   9  16   1   1
0913566091 200
VAR1     1      1.0E1
VAR2     2      5.0E0           1.0     2.0        8.0
VAR3     5      1.0E0         1.0E3
VAR4     6        0.0           5.0    10.0
VAR5     3      5.0E1           2.0     6.0      4.0E2
VAR6     4        1.0          10.0
VAR7     7        5.0         2.5E1     2.0        2.5
VAR8     8        1.0           0.2     2.0        0.8        5.0        1.0
VAR9     9 12        0.5        2.5  0.15  0.15  0.15  0.30  0.30  0.30  0.40  0.40  0.40  0.15  0.15  0.15
```

FIGURE 3·1

SPECIFICATION OF
AN ARBITRARY DISTRIBUTION

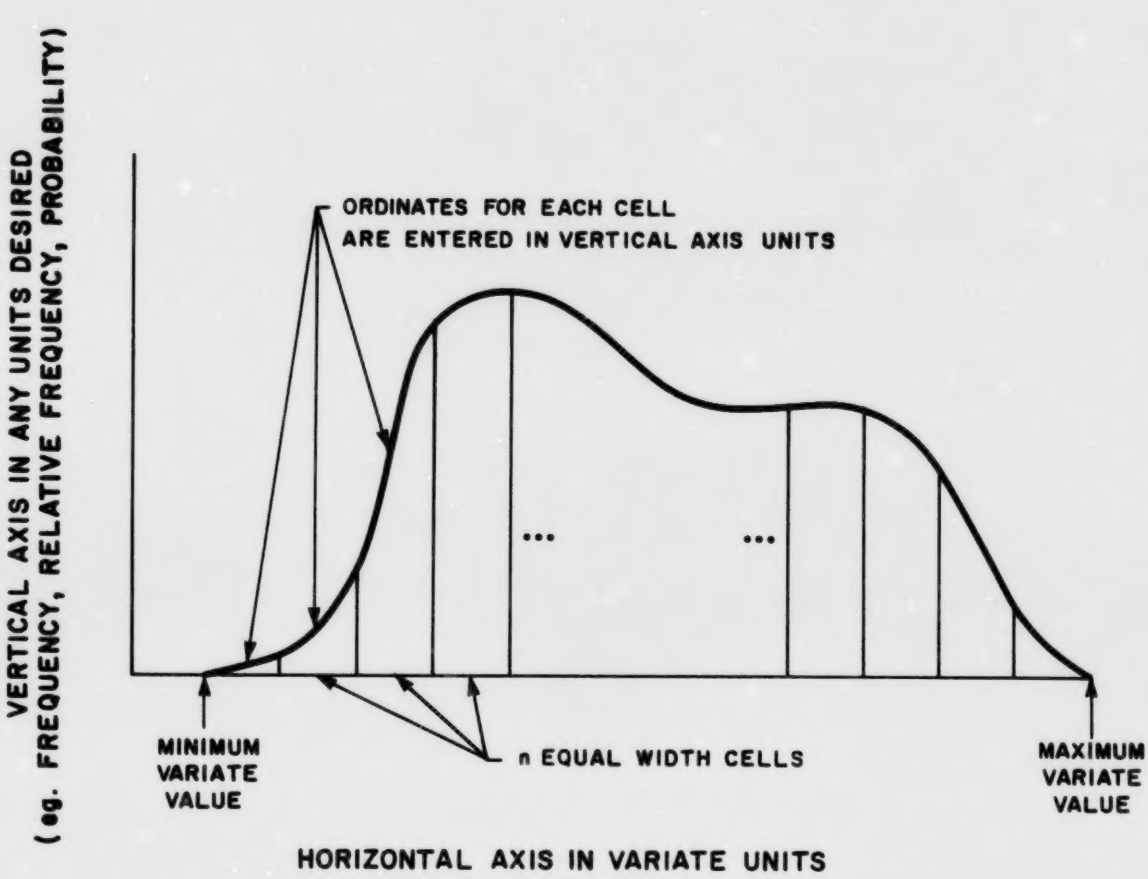
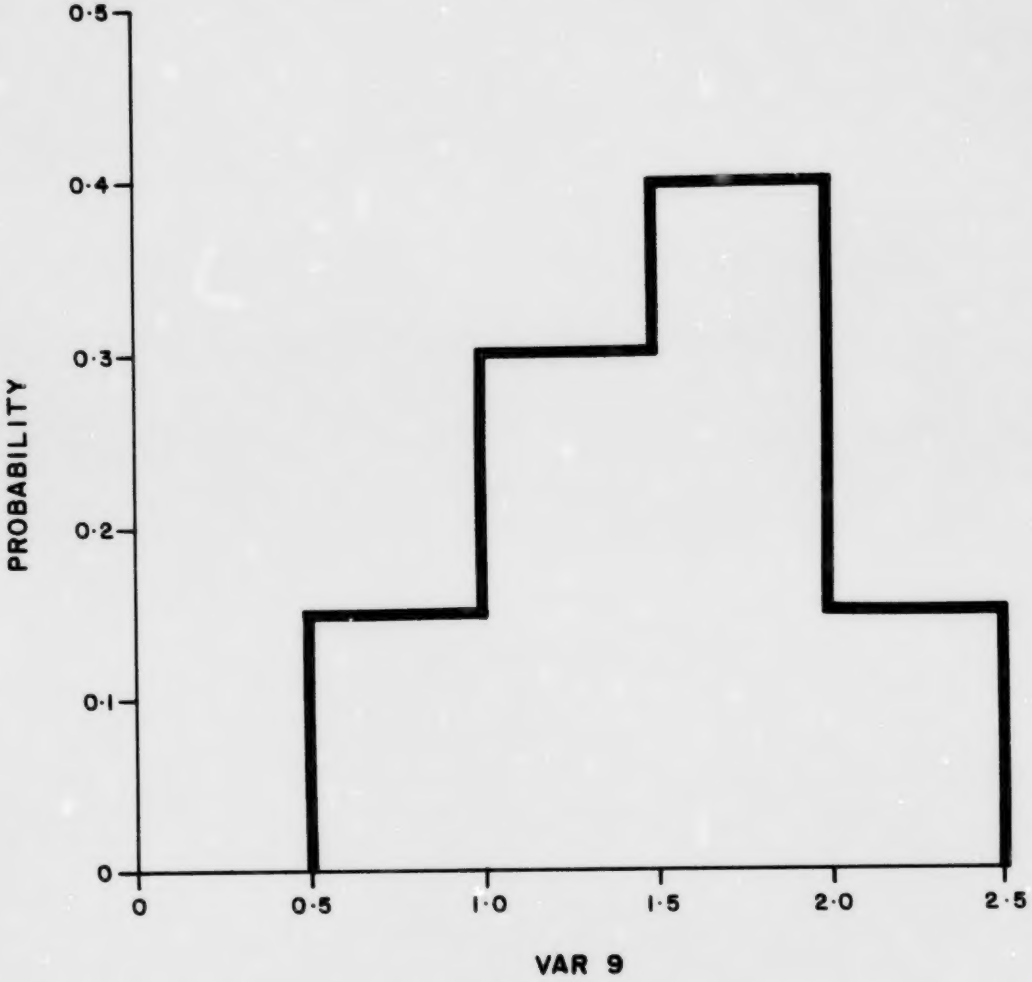

VERTICAL AXIS IN ANY UNITS DESIRED
(eg. FREQUENCY, RELATIVE FREQUENCY, PROBABILITY)
ORDINATES FOR EACH CELL
ARE ENTERED IN VERTICAL AXIS UNITS
...
...
MINIMUM
VARIATE
VALUE
n EQUAL WIDTH CELLS
MAXIMUM
VARIATE
VALUE
HORIZONTAL AXIS IN VARIATE UNITS

FIGURE 3·2

EXAMPLE OF AN ARBITRARY DISTRIBUTION



- a random number seed - in this case the integer 913566091
- a run consisting of 200 probabilistic trials
- 16 cells in the Beta distibution
- 12 cells in the arbitrary distribution.

This example is intended to provide a reference for the implementation of any type of distribution. Further, applied examples of the input files described in this section are shown in Section 4.

In the second input file, the user must supply a parameter name file containing information regarding parameters to be analyzed by RANSIM and to optionally be saved in an output file. Parameters named in this file must be stored in an output vector during program execution. Once again a twelve character name (e.g. EXAMNAME.DAT) must be specified. The format of this file is indicated in Table 3.4. An example of this file is shown in Table 4.5.

RANSIM is an interactive program that prompts for the aforementioned filenames and additional information during program execution. A typical on-line session with RANSIM is detailed on Table 3.5. The messages and prompts from RANSIM are indicated in boldface type while the user responses are in standard type. RANSIM first prompts for the input data file and parameter name file as described earlier. The parameters specified in the parameter name file can optionally be written, for each trial, to an output file. This is accomplished by answering "Y" to the prompt question "Do you wish to save results? (Y/N)" and then specifying a 12 character filename (e.g. OUTPUT11.DAT) to the "Output file name?" prompt. The parameters stored in this file can be either inputs or calculated results written to the output vector during program execution.

Table 3.4

RANSIM Second Input File Characteristics

PARAMETER NAME FILE

Column No.	Format	Parameter Description
Line 1	Title line	
1-40	A40	File title
Line 2		
1-4	I4	number of output parameters i.e. number of records to follows
Line 3 to number of output parameters		
1-6	A6	parameter name
8-48	A40	parameter description

Table 3.5

Example Interactive Session with RANSIM*

```
C>RANSIM

INPUT DATA FILE NAME ? *

EXAMPLE DAT

OPENING FILE: CHECKDOS.DAT

PARAMETER DESCRIPTOR FILE NAME ?

EXAMNAME DAT

DO YOU WISH TO SAVE RESULTS (Y/N)

Y

OUTPUT FILE NAME: ?

OUTPUT11.DAT

OPENING FILE: OUTPUT11.DAT
```

* Boldface type indicates prompts/messages from RANSIM

Once the appropriate information has been accessed by RANSIM, much of it is stored and shared between program modules via named COMMON blocks. Two COMMON blocks referenced in RANSIM are relevant to users:

i) COMMON/INPT/XIN(150)- contains, at any time, a vector of randomly sampled input parameters (from the specified distributions) for transfer to a user-specified subroutine (maximum 150).

ii) COMMON/RESULT/XOUT(150)-contains a vector of parameter values which are to be transferred to RANSIM for analysis and/or storage as output results (maximum 150). These positions in this vector should correspond to the position of the names in the parameter name file.

Further details necessary to implement these concepts are provided using a simple example in Section 4.

4.0 A SIMPLE GENERIC EXAMPLE

4.1 Illustrative Sample

The simple generic example embodies most of the major concepts required to implement a probabilistic model. It uses an extremely simple "simulation model" to illustrate input data file and parameter name file construction and the development of a PAC structure that achieves information transfer to and from the "simulation subroutine". The example illustrates the implementation of a model with a time-varying parameter.

The model has the basic formulation:

$$Y = A+B(1-e^{-CDt})$$

where: Y is the calculated parameter
A, B and C are three inputs parameters
Dt is the time increment (Δt)

Input A has a normal distribution (mean=5,S.D.=1,min=2,max=8) and input B has a triangular distribution (min=10,mode=12,max=16). The input C is a time-varying parameter with a uniform distribution (min=10,max=16). Its value is determined at Dt (Δt) increments of time.

The time increment Δt must also be specified as an input parameter as well as the length of time to be simulated (RYEARS) to allow it to be varied with no programming changes. The capability to store and display the calculated value(Y) for ten equally spaced time intervals (i.e. 11 points in time between 0 and RYEARS inclusive) has been specified a priori.

In general, the stepwise procedure outlined below should be used to develop a system model structure to implement a simulation model:

- Define simulation model interactions
- Define PAC structure
- Define simulation model structure
- Code PAC module
- Code simulation model(s)
- Compile and link RANSIM, PAC and simulation subroutine(s)
- Build input file
- Build parameter name file
- Run

Having defined the simulation model interactions, the next step is to define the PAC structure required to interface with RANSIM and to provide the necessary information transfer to the desired simulation model. This can be done by way of a flow diagram such as that presented in Figure 4.1. It is worth considering the various elements illustrated in more detail before proceeding.

4.2 Details of the PAC Module

The "START" block indicates that the first line of the PAC module must appear in a specific format. This is because RANSIM accesses PAC via the FORTRAN line "CALL PAC (I1)" where I1 is the trial number (a calling argument for the subroutine). Secondly, several common blocks must be defined. The input and result common blocks were mentioned earlier. The order in which the parameters appear in the input common block must correspond to the order in which they appear in the input date file. All variables which appear in any of the common blocks in this example are defined in Table 4.1.

The meta-command $INCLUDE (or $INSERT on some systems) causes any

Table 4.1

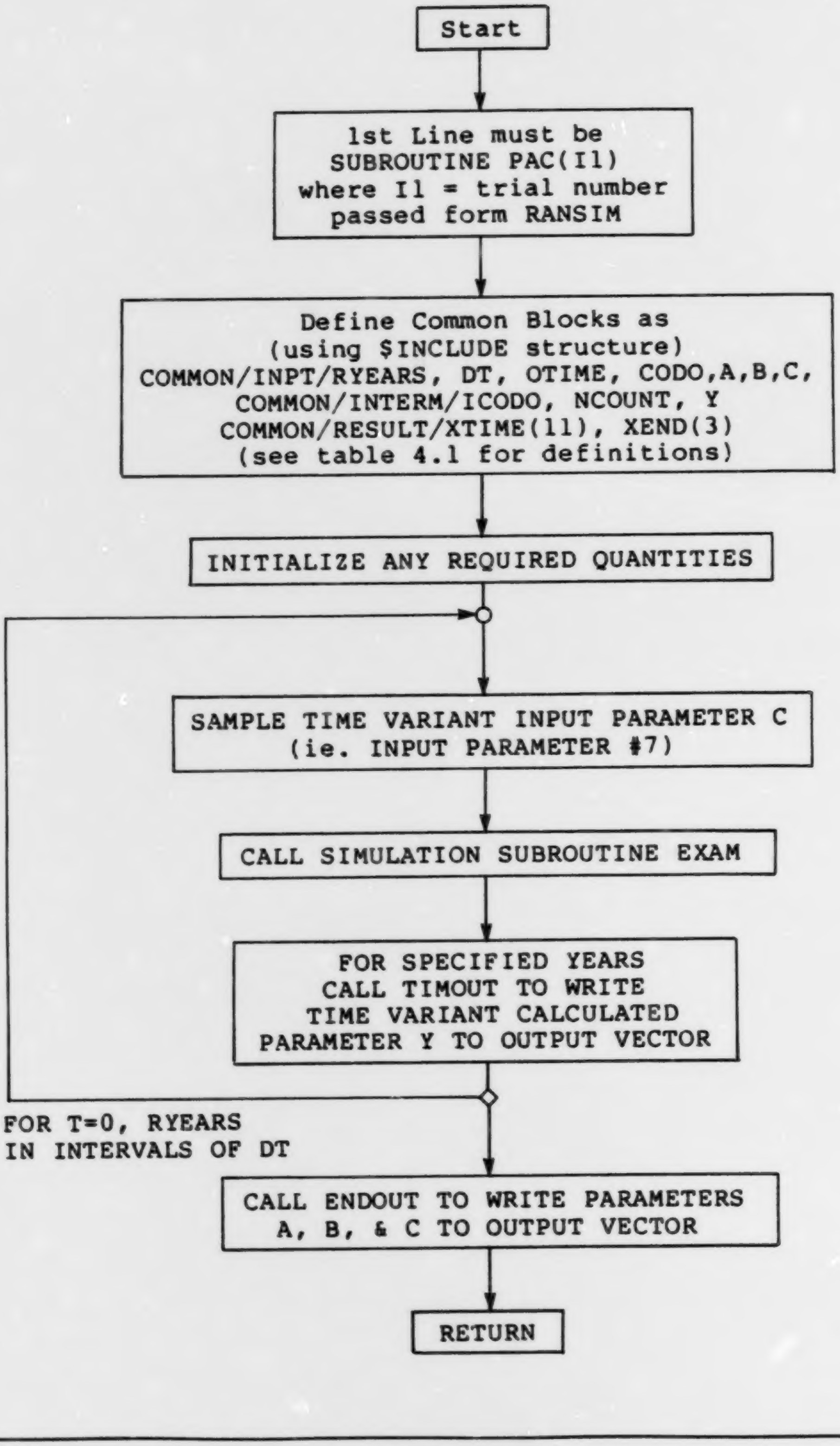
Parameter Definitions for Generic Example

RYEARS*	- length of simulation time period
DT	- simulation time period (Δt)
OTIME*	- time step for which output is to be saved
CODO	- switch to allow display of detailed time variant results to screen (zero=off 1=on)
A,B,C,	- variable input model parameter
ICODO	- integer form of CODO
NCOUNT	- counter based on OTIME to indicate time steps for which results are to be saved
Y	- time-variant model result
XTIME(11)	- vector to store time-variant model results (eg. XTIME(1) = Y at = OTIME etc.) for each trial
XEND(3)	- vector to store values of A, B and C for each trial
I1	- trial number

* Based on the assumptions of this example (RYEARS/OTIME) + 1=11

FIGURE 4·1

SUBROUTINE PAC
FOR SIMPLE EXAMPLE


Start
1st Line must be
SUBROUTINE PAC(I1)
where I1 = trial number
passed form RANSIM
Define Common Blocks as
(using $INCLUDE structure)
COMMON/INPT/RYEARS, DT, OTIME, CODO,A,B,C,
COMMON/INTERM/ICODO, NCOUNT, Y
COMMON/RESULT/XTIME(11), XEND(3)
(see table 4.1 for definitions)
INITIALIZE ANY REQUIRED QUANTITIES
SAMPLE TIME VARIANT INPUT PARAMETER C
(ie. INPUT PARAMETER #7)
CALL SIMULATION SUBROUTINE EXAM
FOR SPECIFIED YEARS
CALL TIMOUT TO WRITE
TIME VARIANT CALCULATED
PARAMETER Y TO OUTPUT VECTOR
FOR T=0, RYEARS
IN INTERVALS OF DT
CALL ENDOUT TO WRITE PARAMETERS
A, B, & C TO OUTPUT VECTOR
RETURN

statements included in the named file to be included and executed at that point in the program. This is an essential feature when there are many large common blocks to be reproduced exactly in several subroutines.

The time-variant nature of this example leads to some of the more complex structures in Figure 4.1. The variables RYEARS AND OTIME must be selected together to ensure that only one specifies the 11 calculated values to be written to the output file. For example one could simulate a 100 year period and save the result for every tenth year(i.e. RYEARS=100,OTIME=10) Note that the restriction of 11 time dependent outputs is arbitrary and arises only because of the original specification. A greater or lesser number of outputs can be obtained with minor programming changes. The time step Δt can be arbitrarily selected and can be easily changed in the input data file.

The common block INTERM, is constructed to allow information which is transferred only between PAC and the user-specified model to be accessed easily.

Once the common blocks are defined and any necessary quantities are intialized, a time loop is defined. Instructions within the time loop will be carried out for every time from zero to RYEARS in steps of Dt (Δt). The first instruction involves selecting a value for the time-variant input parameter C (parameter #7 in the INPT list). As will be seen shortly, this can be accomplished by the FORTRAN command:

```
CALL CHOOSE (7)
```

where CHOOSE indicates a subroutine in RANSIM which performs the sampling and the parameter (7) indicates that CHOOSE is to select an appropiate value for the seventh input parameter as defined in the input data file discussed in Section 4.7.

Once the value for the parameter C has been selected, all model input parameters (A,B,C, and Dt) have been defined and the simulation subroutine EXAM (the name of the example model) can be called. It will perform the calculations specified and return a value for Y for that time interval. If the time interval corresponds to OTIME, the Y value will be saved, otherwise, the time loop instructions will be repeated. Once the time loop is completed at t=RYEARS the input parameters A,B, and C will be saved. The value saved for C will be that for the final time interval.

4.3 Structure of the Simulation Module

Having defined the PAC structure appropriate for this example, the structure of the simulation model itself must be outlined. For EXAM, this is a relatively trivial task but it is presented here for completeness. The simulation model structure is shown in Figure 4.2. As indicated, the common block definition is identical to that for the PAC module. The presence of the "switch"(ICODO) allows detailed results to be displayed. It is recommended that many of these detailed printout codes and write statements be kept to facilitate error tracing and quality assurance. The other operations indicated are fairly self-explanatory.

4.4 Coding of PAC and EXAM

While the information presented in the preceding sections should be sufficient to perform a probabilistic analysis of the example, the implementation of the PAC subroutines and subroutine EXAM is presented in Table 4.2 for completeness. Most of the code uses conventional Fortran commands and structure. One unusual feature

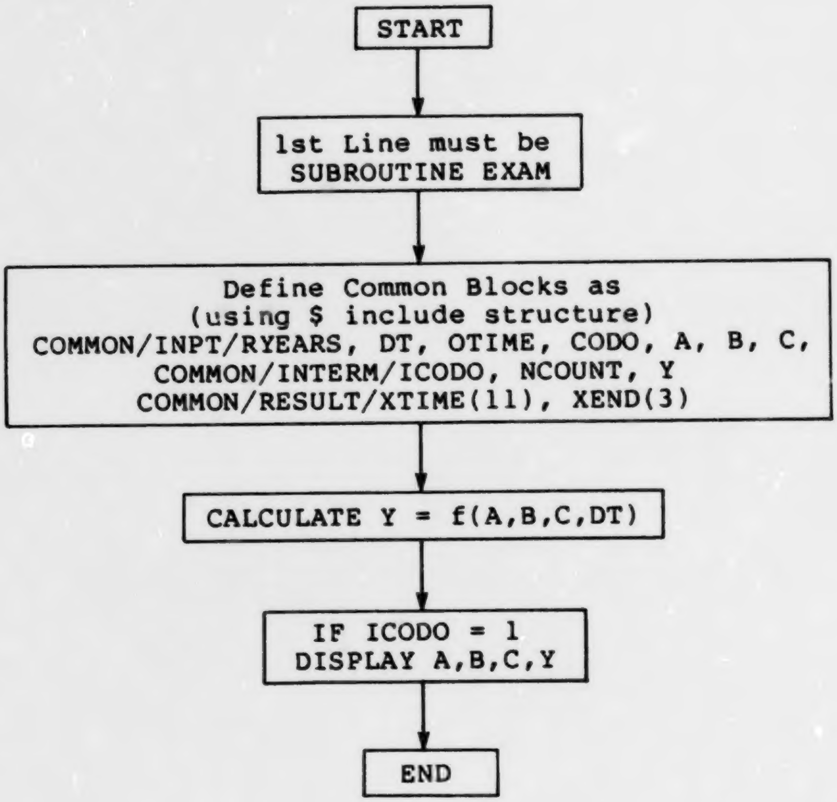
Table 4.2

PAC, ENDOUT, TIMOUT, and EXAM Subroutines

```
$DEBUG
      SUBROUTINE PAC(I1)
      INTEGER*2 I1
$INCLUDE:'B:STUFF.INC'
      WRITE(*,20)I1
 20   FORMAT(1X,'PAC CALLED ON TRIAL NUMBER',I5)
C  INITIALIZE QUANTITIES AT BEGINNING OF SIMULATION
      IF(I1.EQ.1) THEN
        NYEARS = INT(RYEARS+DT)
        NDT = INT(DT)
        ICODO=INT(CODO)
      ENDIF
      COUNT = 0.0
C   BEGIN TIME LOOP
      DO 1 IYR = NDT,NYEARS,NDT
C   DETERMINE TIME T
        XIYR=FLOAT(IYR)
        T=XIYR-DT
C   SAMPLE TIME DEPENDENT INPUT
        CALL CHOOSE(7)
C   CALL SIMULATION SUBROUTINE
        CALL EXAM
C   SAVE TIME DEPENDENT OUTPUT AT APPROPRIATE TIME INTERVAL
        IF(COUNT.EQ.(T/OTIME))THEN
        NCOUNT=INT(COUNT+1.0)
        CALL TIMOUT
        COUNT=COUNT+1.0
       ENDIF
 1     CONTINUE
C   SAVE INPUTS
       CALL ENDOUT
       RETURN
      END
C
      SUBROUTINE ENDOUT
C   SAVES INPUTS
$INCLUDE:'B:STUFF.INC'
      XEND(1)=A
      XEND(2)=B
      XEND(3)=C
        RETURN
       END
C
       SUBROUTINE TIMOUT
C   SAVES TIME DEPENDENT CALCULATIONS
$INCLUDE:'B:STUFF.INC'
       XTIME(NCOUNT)=Y
        RETURN
       END
C
      SUBROUTINE EXAM
C   EXAMPLE SIMULATION SUBROUTINE
$INCLUDE:'B:STUFF.INC'
      Y=A+B*(1-EXP(-C*DT))
      IF(ICODO.EQ.1)WRITE(*,10)Y,A,B,C
 10   FORMAT(1X,'Y=',E13.6,1X,'A=',E13.6,1X,'B=',E13.6,1X,'C=',E13.6)
      RETURN
      END
```

FIGURE 4·2

SUBROUTINE EXAM


START
1st Line must be
SUBROUTINE EXAM
Define Common Blocks as
(using $ include structure)
COMMON/INPT/RYEARS, DT, OTIME, CODO, A, B, C,
COMMON/INTERM/ICODO, NCOUNT, Y
COMMON/RESULT/XTIME(11), XEND(3)
CALCULATE Y = f(A,B,C,DT)
IF ICODO = 1
DISPLAY A,B,C,Y
END

is the file "B:STUFF.INC" in line 3 of PAC. Table 4.3 is a listing of the file STUFF.INC. The values stored in COMMON/RESULT/ for each trail will be XTIME (1 to 11) (i.e. Y (O) to Y (RYEARS)), followed by XEND (1 to 3) (i.e. A, B and C). More time-variant output values can be saved by changing the dimension of XTIME( ) and the parameter name file (described in Section 4.6).

4.5 Compile and Link

Once the PAC and EXAM modules are coded, RANSIM, PAC and EXAM can be compiled using an appropriate FORTRAN compiler. When a "clean compile" is achieved (i.e. no fatal programming errors), the object modules (discussed in Section 4.7) should be linked to produce an executable file.

4.6 Input Data File and Parameter Name File

The input data file required by this program is presented in Table 4.4. The comments in boldface type on the table would not be present in the actual data set, and are included for explanatory purposes only. The data file corresponds to the structure outlined in Table 3.1 and the order of the parameters corresponds to that in COMMON (INPT). The data file presented is for 10 trials with 100 years per trial, a Δt of 10 years, and saving the result at every tenth year. The detailed print switch has been turned "on" (i.e. ICODO=1.0).

The second data file, containing variable names and descriptions of the parameters stored in COMMON/RESULT/ for analysis and/or storage is presented in Table 4.5. Again, boldface comments are explanatory and would not appear in the actual file. As indicated, all values of the time-variant calculated results will be displayed before the input parameters A, B, and C. This

Table 4.3

STUFF.INC File

```
C  INPUT COMMON BLOCK
      COMMON/INPT/RYEARS,DT,OTIME,CODO,A,B,C
C  COMMON BLOCK FOR PRINT CODES
      COMMON/INTERM/ICODO,NCOUNT
C  COMMON BLOCK FOR INTERMEDIATE VALUES
      COMMON/INTERM/Y
C  OUTPUT COMMON BLOCK
      COMMON/RESULT/XTIME(11),XEND(3)
```

TABLE 4.4
INPUT PARAMETERS

```
GENERIC EXAMPLE DATASET                LINE 1 TITLE

LINE 2  NUMBER OF BETA   NUMBER OF  NUMBER OF CELLS IN  NUMBER OF               NUMBER OF
        DISTRIBUTIONS    INPUTS     BETA DISTRIBUTION   CORRELATED PARAMETERS   ARBITRARY DISTRIBUTIONS

  0  7  0  0  0

LINE 3  RANDOM NUMBER SEED   NUMBER OF TRIALS

0913566091  10
RYEARS   1    1.00E2  ------------------------------  NUMBER OF YEARS IN SIMULATION
DT       1    1.0E01  ------------------------------  TIME INCREMENT
OTIME    1    1.0E01  ------------------------------  OUTPUT EVERY OTIME YEARS
CODO     1    0.0     ------------------------------  DETAILED PRINTOUT FROM  SIMULATION SUBROUTINE
A        2    5.00E0    1.0E00    3.0E00    8.0E00 -]
B        6    1.0E01    1.2E01    1.6E01    0.0    -]  INPUT PARAMETERS
C        4    5.0E-1    1.0E00    0.0       0.0    -]
```

* Boldface type indicates explanatory comments not present in actual data file

TABLE 4.5

OUTPUT PARAMETER NAME FILE*

LINE 1 TITLE

```
OUTPUT PARAMETERS
```

LINE 2 NUMBER OF OUTPUT PARAMETERS

```
  14
Y0    Y AT T=0                    :
Y10   Y AT T=10                   :
Y20   Y AT T=20                   :
Y30   Y AT T=30                   :
Y40   Y AT T=40                   :
Y50   Y AT T=50                   :
Y60   Y AT T=60                   :
Y70   Y AT T=70                   :
Y80   Y AT T=80                   :
Y90   Y AT T=90                   :
Y100  Y AT T=100                  :
A     INPUT PARAMETER A           :
B     INPUT PARAMETER B           :
C     INPUT PARAMETER C           :
```

OUTPUT PARAMETER NAMES AND DESCRIPTIONS

* Boldface type indicates explanatory comments which would not appear in actual data file

corresponds to the order of their appearance in the COMMON/RESULT vector.

4.7 Running The Program And Sample Results

To run the example, all subroutines must be compiled to produce object modules and then linked with the RANSIM module to produce an executable program. Detailed results (i.e. with CODO=1.0) generated by such a program are presented in Table 4.6 for three trials. Table 4.7 shows the summary statistics displayed to the screen at the conclusion of the run. Note that the summary statistics displayed to the screen are also saved in a file labelled "STATS" . The program also saves the random number seed for the "current" iteration in a file called "RANSEED". This is a useful feature, if for some reason, program execution is terminated by a run-time error. The error could be replicated quite simply by re-running the program using the random number seed indicated in the file. If no errors occur, the file simply contains the random number seed for the final trial.

TABLE 4.6

PRINTOUT OF THREE TRIALS FOR THE EXAMS MODEL

```
PAC CALLED ON TRIAL NUMBER     1
Y=  .170840E+02 A=  .488308E+01 B=  .122221E+02 C=  .636032E+00
Y=  .170915E+02 A=  .488308E+01 B=  .122221E+02 C=  .679555E+00
Y=  .170974E+02 A=  .488308E+01 B=  .122221E+02 C=  .736790E+00
Y=  .171045E+02 A=  .488308E+01 B=  .122221E+02 C=  .976384E+00
Y=  .171026E+02 A=  .488308E+01 B=  .122221E+02 C=  .845972E+00
Y=  .170947E+02 A=  .488308E+01 B=  .122221E+02 C=  .706320E+00
Y=  .171016E+02 A=  .488308E+01 B=  .122221E+02 C=  .814585E+00
Y=  .170627E+02 A=  .488308E+01 B=  .122221E+02 C=  .566185E+00
Y=  .170401E+02 A=  .488308E+01 B=  .122221E+02 C=  .523635E+00
Y=  .170889E+02 A=  .488308E+01 B=  .122221E+02 C=  .661984E+00
Y=  .170929E+02 A=  .488308E+01 B=  .122221E+02 C=  .690141E+00
PAC CALLED ON TRIAL NUMBER     2
Y=  .161699E+02 A=  .519852E+01 B=  .109723E+02 C=  .931744E+00
Y=  .161328E+02 A=  .519852E+01 B=  .109723E+02 C=  .566512E+00
Y=  .161622E+02 A=  .519852E+01 B=  .109723E+02 C=  .714000E+00
Y=  .161685E+02 A=  .519852E+01 B=  .109723E+02 C=  .842304E+00
Y=  .161522E+02 A=  .519852E+01 B=  .109723E+02 C=  .637652E+00
Y=  .161699E+02 A=  .519852E+01 B=  .109723E+02 C=  .937797E+00
Y=  .161245E+02 A=  .519852E+01 B=  .109723E+02 C=  .546634E+00
Y=  .161516E+02 A=  .519852E+01 B=  .109723E+02 C=  .634408E+00
Y=  .161671E+02 A=  .519852E+01 B=  .109723E+02 C=  .797057E+00
Y=  .161481E+02 A=  .519852E+01 B=  .109723E+02 C=  .617635E+00
Y=  .161611E+02 A=  .519852E+01 B=  .109723E+02 C=  .702726E+00
PAC CALLED ON TRIAL NUMBER     3
Y=  .165870E+02 A=  .441171E+01 B=  .121800E+02 C=  .787328E+00
Y=  .165474E+02 A=  .441171E+01 B=  .121800E+02 C=  .561758E+00
Y=  .165601E+02 A=  .441171E+01 B=  .121800E+02 C=  .595489E+00
Y=  .165868E+02 A=  .441171E+01 B=  .121800E+02 C=  .781642E+00
Y=  .165911E+02 A=  .441171E+01 B=  .121800E+02 C=  .995435E+00
Y=  .165911E+02 A=  .441171E+01 B=  .121800E+02 C=  .991004E+00
Y=  .165682E+02 A=  .441171E+01 B=  .121800E+02 C=  .625233E+00
Y=  .165842E+02 A=  .441171E+01 B=  .121800E+02 C=  .739332E+00
Y=  .165911E+02 A=  .441171E+01 B=  .121800E+02 C=  .995376E+00
Y=  .165850E+02 A=  .441171E+01 B=  .121800E+02 C=  .751501E+00
Y=  .165483E+02 A=  .441171E+01 B=  .121800E+02 C=  .563727E+00
```

Table 4.7

Summary Statistics from 10 Trials

VARIABLE	MEAN	STD DEV	SKEWNESS	KURTOSIS	GEOM MEAN	GEOM SD	MINIMUM	MAXIMUM
Y0	.1712E+02	.1263E+01	-.6138E+00	.2867E+01	.1708E+02	.1079E+01	.1452E+02	.1882E+02
Y10	.1713E+02	.1262E+01	-.5552E+00	.2724E+01	.1708E+02	.1078E+01	.1457E+02	.1881E+02
Y20	.1712E+02	.1255E+01	-.5342E+00	.2821E+01	.1708E+02	.1078E+01	.1455E+02	.1881E+02
Y30	.1713E+02	.1256E+01	-.6327E+00	.2833E+01	.1708E+02	.1078E+01	.1453E+02	.1875E+02
Y40	.1712E+02	.1243E+01	-.6354E+00	.2831E+01	.1708E+02	.1077E+01	.1456E+02	.1874E+02
Y50	.1713E+02	.1251E+01	-.5925E+00	.2806E+01	.1709E+02	.1078E+01	.1457E+02	.1881E+02
Y60	.1712E+02	.1255E+01	-.5655E+00	.2757E+01	.1708E+02	.1078E+01	.1458E+02	.1882E+02
Y70	.1713E+02	.1256E+01	-.5703E+00	.2773E+01	.1709E+02	.1078E+01	.1457E+02	.1881E+02
Y80	.1713E+02	.1251E+01	-.5879E+00	.2784E+01	.1709E+02	.1078E+01	.1457E+02	.1878E+02
Y90	.1713E+02	.1254E+01	-.5916E+00	.2790E+01	.1709E+02	.1078E+01	.1456E+02	.1880E+02
Y100	.1713E+02	.1256E+01	-.5719E+00	.2766E+01	.1709E+02	.1078E+01	.1457E+02	.1881E+02
A	.4683E+01	.6825E+00	.2226E+00	.2141E+01	.4639E+01	.1157E+01	.3774E+01	.5917E+01
B	.1246E+02	.1198E+01	.3281E+00	.2227E+01	.1241E+02	.1100E+01	.1080E+02	.1454E+02
C	.7471E+00	.1217E+00	-.7774E-01	.1839E+01	.7379E+00	.1182E+01	.5637E+00	.9238E+00

5.0 THE URANIUM TAILINGS ASSESSMENT PROGRAM

This section discusses the input files and PAC structure required to run the probabilistic Uranium Tailings Assessment Program (UTAP). The conceptual structure of UTAP is illustrated in Figure 5.1. Program listings of the simulation subroutines are presented in Appendix B while detailed mathematical descriptions and variable dictionaries for the simulation subroutines are listed in Appendices A to F.

The input data file conforming to the format described in Tables 3.1 and 3.2 is shown in Table 5.1. Print codes (those variables that have names beginning with COD) to turn the detailed printout on or off in the simulation subroutines are used frequently for error checking and debugging. The parameter name file detailing the parameters to be analyzed and optionally stored by RANSIM on disk is shown in Table 5.2. This file conforms to the format description presented in Table 3.4.

Source listings of subroutine PAC, TIMOUT, and ENDOUT are presented in Table 5.3. The file STUFF.INC used in the $INCLUDE statement in the above three subroutines and in the simulation subroutines is listed in Table 5.4. A dictionary of the variables used in PAC, TIMOUT and ENDOUT is shown in Table 5.5.

A maximum of 150 parameters per trial can be written to the output array for analysis and optional disk storage and statistics. The model developed to characterize the reference tailings example discussed in Volume 1 stores 62 input parameters and eight time-dependent parameters (at 11 evenly spaced time intervals from 0 to the end of the simulation) in this array. The interval between storage times is RYEARS (see Table 5.5) divided by 10 (which equals OTIME).

FIGURE 5·1

CONCEPTUAL STRUCTURE OF THE URANIUM TAILINGS ASSESSMENT PROGRAM

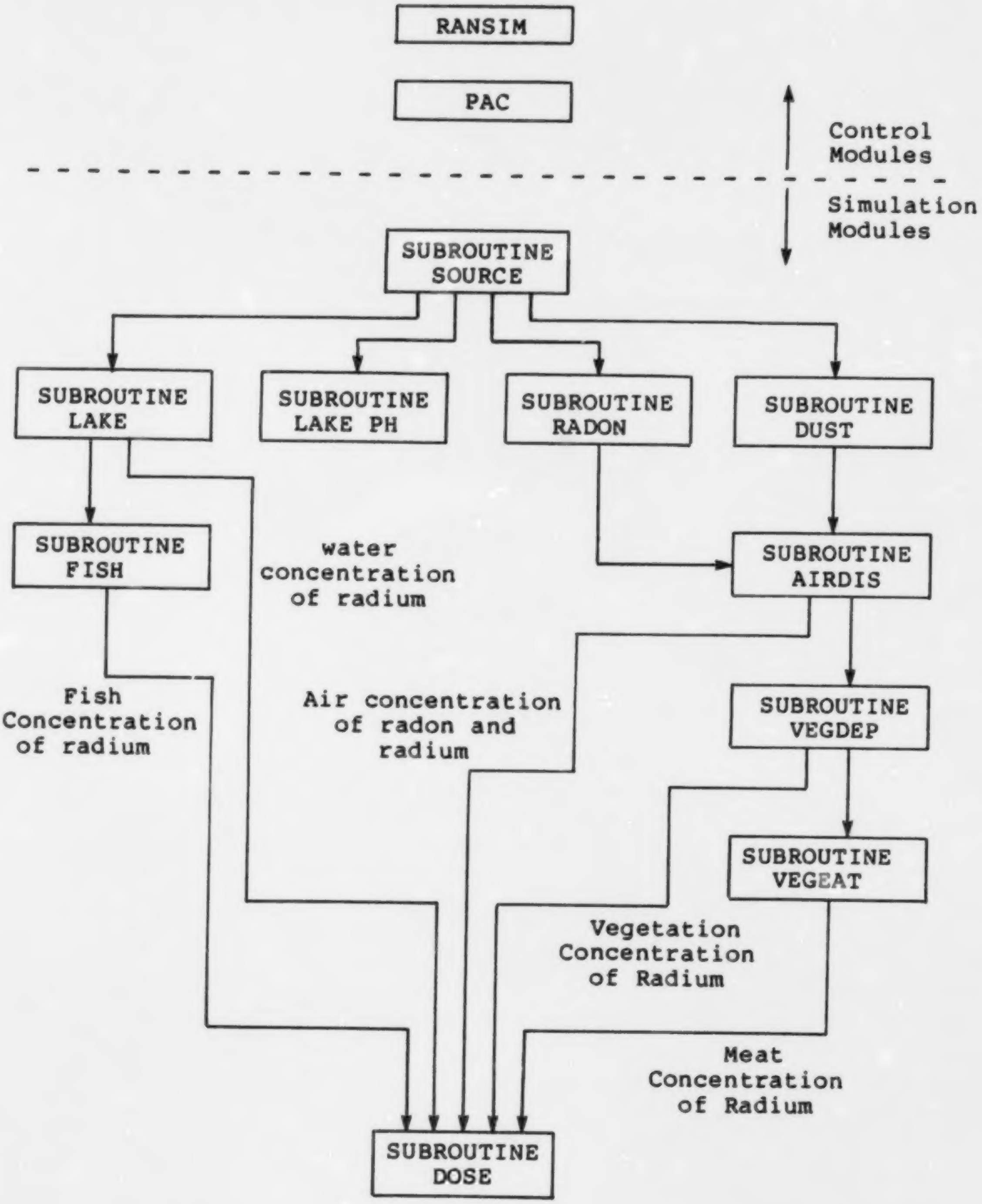


Table 5.1

Input Parameters That Will Fill COMMON/INPT/

```
PROBABILISTIC ANALYSIS:URANIUM TAILINGS
    0  97  0   0    0
0913566091   20
FWIND1    2      3.5E1       3.0E-1     3.41E01     3.59E01
FWIND2    2      .1E00       8.0E-4     9.73E-2     1.03E-1
FPREC     2     1.23E2       8.0E-1     1.21E02      1.25E2
UWIND     2     5.3E00       2.2E-2      5.2E00      5.4E00
RYEARS    1      5.0E2
DT        1     1.0E01
OTIME     1     5.0E01
CODS      1     1.0E00
CODL      1     0.0E00
CODPH     1     0.0E00
CODF      1     0.0E00
CODR      1     0.0E00
CODD      1     0.0E00
CODA      1     0.0E00
CODV1     1     0.0E00
CODV2     1     0.0E00
CODO      1     0.0E00
CODU      1     0.0E00
SULPRD    3     1.4E-6       1.73E0      2.7E-7      7.2E-6
RHOTAL    2      1.4E6        0.1E6      1.1E06      1.7E06
O2CONC    1       9.04
O2DIFF    3      7.0E1       1.35E0      2.8E01      1.7E02
KPRIME    3     5.0E-2          2.0      6.2E-3      4.0E-1
KSUBSO    1      2.5E1          0.0         0.0         0.0
NPRCP     2     4.0E-1       4.0E-3      3.9E-1      4.1E-1
FACTCO    6     2.0E-1       3.0E-1      5.0E-1         0.0
DISTCO    4     1.0E-2       1.0E-1
DUNSAT    1        1.0          1.5
DMIX      1        1.0          1.5
KSUBDR    3    80.0E-6       3.5E00      1.9E-6      3.4E-3
RAINGP    6     5.0E-2       2.0E-1      6.0E-1
TVOID     6     3.5E-1       4.0E-1      4.5E-1
CONCRA    6        5.0        1.1E1       1.5E1
FRACGP    4     4.0E-2       6.0E-2         0.0
PYRITE    6     5.5E-2       7.3E02      9.0E02
UPERTH    4     2.0E00       4.0E00         0.0
UO2TON    6     1.0E00       2.0E00      3.0E00
RAINTH    6     8.5E-1       9.0E-1      9.5E-1
THDISS    4     1.0E-2       3.0E-2         0.0
THFRAC    6        0.2       3.0E-1         0.4
XK1       3     6.3E-1     1.85E00     1.80E-1     2.15E00
XK3       3     2.2E-2     1.20E00     1.86E-2     2.66E-2
BETA      3     1.6E-1     1.35E00      8.9E-2     3.03E-1
ALAKE     1     1.6E06
DLAKE     1     1.0E01
AWAT      1     2.0E07
ATAIL     1     5.0E05
```

Table 5.1 continued

ACTCO	1	0.8E00			
BF	4	1.50E-7	1.50E-6		
MPCH2O	6	5.0E00	1.5E01	2.5E01	
RNLAM	1	2.1E-6			
EMANAT	6	2.0E-2	2.0E-1	9.0E-1	
SILT	1	.3E2			
AEXPT	1	5.0E5			
HSRC	4	1.0E00	3.0E00		
XDIST	1	3.6E03			
VDISP	3	6.0E01	1.73E0	1.16E1	3.11E2
FDEPSF1	4	8.0E-1	1.0E00		
FDEPSF2	2	4.7E-1	3.0E-1	1.0E-2	1.0E00
FDEPSF3	4	1.0E-1	3.0E-1		
FDEPSF4	4	1.0E-1	3.0E-1		
FDEPED1	4	8.0E-1	1.0E00		
FDEPED2	4	8.0E-1	1.0E00		
FDEPED3	4	8.0E-1	1.0E00		
FDEPED4	4	5.0E-2	1.5E-1		
DUREXP1	4	9.0E07	1.1E08		
DUREXP2	4	2.7E06	3.1E06		
DUREXP3	4	4.7E06	5.6E06		
DUREXP4	4	4.7E06	5.6E06		
DEPVEL	4	3.0E-3	3.0E-2		
VEGLAM1	4	2.0E-9	4.0E-9		
VEGLAM2	4	4.0E-7	6.0E-7		
VEGLAM3	4	4.0E-7	6.0E-7		
VEGLAM4	4	4.0E-7	6.0E-7		
YDEN1	4	4.0E02	6.0E02		
YDEN2	4	5.0E02	1.0E03		
YDEN3	4	1.5E03	2.5E03		
YDEN4	4	1.5E03	2.5E03		
FBI	4	.30E-9	2.5E-9		
FAUT1	6	0.50E-1	0.2E00	.50E00	
FAUT2	6	0.50E-1	0.2E00	.50E00	
FIR	4	2.25E03	2.75E03		
DCFING	1	3.200E-7			
VEGINA	1	1.570E05			
VEGINB	1	7.0E04			
FACTA	4	5.0E-1	1.0E0		
FACTB	4	5.0E-1	1.0E0		
FLOCA	6	.1E00	.30E0	.6E00	
XMEATI	1	7.9E04			
FMEAT	4	0.0E00	1.0E0		
WATIN	1	.6E00			
FISHIN	4	4.0E03	1.2E04		
FFISH	1	1.0E00			
DCFINH	1	2.6E-6			
XINHIN	1	8.4E03			
OF	6	4.0E-1	6.0E-1	9.0E-1	
F	6	2.0E-1	4.0E-1	7.0E-1	

Table 5.2

Parameter Name File For Parameters That Will Fill COMMON/RESULT/

```
 150
DV0     DOSE AT T=0 VEGETATION                 :
DV50    DOSE AT T=50 VEGETATION                :
DV100   DOSE AT T=100 VEGETATION               :
DV150   DOSE AT T=150 VEGETATION               :
DV200   DOSE AT T=200 VEGETATION               :
DV250   DOSE AT T=250 VEGETATION               :
DV300   DOSE AT T=300 VEGETATION               :
DV350   DOSE AT T=350 VEGETATION               :
DV400   DOSE AT T=400 VEGETATION               :
DV450   DOSE AT T=450 VEGETATION               :
DV500   DOSE AT T=500 VEGETATION               :
DM0     DOSE AT T=0 MEAT                       :
DM50    DOSE AT T=50 MEAT                      :
DM100   DOSE AT T=100 MEAT                     :
DM150   DOSE AT T=150 MEAT                     :
DM200   DOSE AT T=200 MEAT                     :
DM250   DOSE AT T=250 MEAT                     :
DM300   DOSE AT T=300 MEAT                     :
DM350   DOSE AT T=350 MEAT                     :
DM400   DOSE AT T=400 MEAT                     :
DM450   DOSE AT T=450 MEAT                     :
DM500   DOSE AT T=500 MEAT                     :
DW0     DOSE AT T=0 DRINKING WATER             :
DW50    DOSE AT T=50 DRINKING WATER            :
DW100   DOSE AT T=100 DRINKING WATER           :
DW150   DOSE AT T=150 DRINKING WATER           :
DW200   DOSE AT T=200 DRINKING WATER           :
DW250   DOSE AT T=250 DRINKING WATER           :
DW300   DOSE AT T=300 DRINKING WATER           :
DW350   DOSE AT T=350 DRINKING WATER           :
DW400   DOSE AT T=400 DRINKING WATER           :
DW450   DOSE AT T=450 DRINKING WATER           :
DW500   DOSE AT T=500 DRINKING WATER           :
DF0     DOSE AT T=0 FISH                       :
DF50    DOSE AT T=50 FISH                      :
DF100   DOSE AT T=100 FISH                     :
DF150   DOSE AT T=150 FISH                     :
DF200   DOSE AT T=200 FISH                     :
DF250   DOSE AT T=250 FISH                     :
DF300   DOSE AT T=300 FISH                     :
DF350   DOSE AT T=350 FISH                     :
DF400   DOSE AT T=400 FISH                     :
DF450   DOSE AT T=450 FISH                     :
DF500   DOSE AT T=500 FISH                     :
DD0     DOSE AT T=0 DUST INHALATION            :
DD50    DOSE AT T=50 DUST INHALATION           :
DD100   DOSE AT T=100 DUST INHALATION          :
DD150   DOSE AT T=150 DUST INHALATION          :
DD200   DOSE AT T=200 DUST INHALATION          :
```

Table 5.2 continued

DD250	DOSE AT T=250 DUST INHALATION	:
DD300	DOSE AT T=300 DUST INHALATION	:
DD350	DOSE AT T=350 DUST INHALATION	:
DD400	DOSE AT T=400 DUST INHALATION	:
DD450	DOSE AT T=450 DUST INHALATION	:
DD500	DOSE AT T=500 DUST INHALATION	:
DR0	DOSE AT T=0 RADON INHALATION	:
DR50	DOSE AT T=50 RADON INHALATION	:
DR100	DOSE AT T=100 RADON INHALATION	:
DR150	DOSE AT T=150 RADON INHALATION	:
DR200	DOSE AT T=200 RADON INHALATION	:
DR250	DOSE AT T=250 RADON INHALATION	:
DR300	DOSE AT T=300 RADON INHALATION	:
DR350	DOSE AT T=350 RADON INHALATION	:
DR400	DOSE AT T=400 RADON INHALATION	:
DR450	DOSE AT T=450 RADON INHALATION	:
DR500	DOSE AT T=500 RADON INHALATION	:
TD0	TOTAL DOSE AT T=0	:
TD50	TOTAL DOSE AT T = 50	:
TD100	TOTAL DOSE AT T=100	:
TD150	TOTAL DOSE AT T=150	:
TD200	TOTAL DOSE AT T=200	:
TD250	TOTAL DOSE AT T=250	:
TD300	TOTAL DOSE AT T=300	:
TD350	TOTAL DOSE AT T=350	:
TD400	TOTAL DOSE AT T=400	:
TD450	TOTAL DOSE AT T=450	:
TD500	TOTAL DOSE AT T=500	:
PH0	PH AT T=0	:
PH50	PH AT T=50	:
PH100	PH AT T=100	:
PH150	PH AT T=150	:
PH200	PH AT T=200	:
PH250	PH AT T=250	:
PH300	PH AT T=300	:
PH350	PH AT T=350	:
PH400	PH AT T=400	:
PH450	PH AT T=450	:
PH500	PH AT T=500	:
SULPRD		:
RHOTAL		:
O2DIFF		:
KPRIME		:
NPRCP		:
FACTCO		:
DISTCO		:
DUNSAT		:
KSUBDR		:
TVOID		:
FRACGP		:
PYRITE		:
UPERTH		:
UO2TON		:

Table 5.2 continued

RAINTH :
THDISS :
RAINGP :
CONCRA :
THFRAC :
XK1 :
XK3 :
BETA :
BF :
MPCH2O :
EMANAT :
FPREC :
FWIND1 :
FWIND2 :
HSRC :
UWIND :
VDISP :
DEPVEL :
FDEPSF1 :
FDEPSF2 :
FDEPSF3 :
FDEPSF4 :
FDEPED1 :
FDEPED2 :
FDEPED3 :
FDEPED4 :
DUREXP1 :
DUREXP2 :
DUREXP3 :
DUREXP4 :
VEGLAM1 :
VEGLAM2 :
VEGLAM3 :
VEGLAM4 :
YDEN1 :
YDEN2 :
YDEN3 :
YDEN4 :
FAUT1 :
FAUT2 :
FBI :
FACTA :
FACTB :
FLOCA :
FMEAT :
FISHIN :
OF :
F :

TABLE 5.3

Source Code For PAC, TIMOUT, and ENDOUT

```
$DEBUG
      SUBROUTINE PAC(I1)
      INTEGER*2 I1
$INCLUDE:'B:STUFF.INC'
      WRITE(*,20)I1
 20   FORMAT(1X,'PAC CALLED ON TRIAL NUMBER',I5)
      IF(I1.EQ.1) THEN
        NYEARS = INT(RYEARS+DT)
        NDT = INT(DT)
        ICODS=INT(CODS)
        ICODPH=INT(CODPH)
        ICODL=INT(CODL)
        ICODF=INT(CODF)
        ICODR=INT(CODR)
        ICODD=INT(CODD)
        ICODA=INT(CODA)
        ICODV1=INT(CODV1)
        ICODV2=INT(CODV2)
        ICODO=INT(CODO)
      ENDIF
      XI1=FLOAT(I1)
      COUNT = 0.0
      DO 1 IYR = NDT,NYEARS,NDT
        XIYR=FLOAT(IYR)
        T=XIYR-DT
        CALL SOURCE
        CALL LAKE
        CALL LAKEPH
        CALL FISH
        CALL RADON
        CALL DUST
        CALL AIRDIS
        CALL VEGDEP
        CALL VEGEAT
        CALL DOSEIN
        IF(COUNT.EQ.(T/OTIME))THEN
        NCOUNT=INT(COUNT+1.0)
        CALL TIMOUT
        COUNT=COUNT+1.0
      ENDIF
 1    CONTINUE
      CALL ENDOUT
      RETURN
      END
C
      SUBROUTINE ENDOUT
$INCLUDE:'B:STUFF.INC'
      XEND(1)=SULPRD
```

Table 5.3 continued

```
      XEND(2)=RHOTAL
      XEND(3)=O2DIFF
      XEND(4)=KPRIME
      XEND(5)=NPRCP
      XEND(6)=FACTCO
      XEND(7)=DISTCO
      XEND(8)=DUNSAT
      XEND(9)=KSUBDR
      XEND(10)=TVOID
      XEND(11)=FRACGP
      XEND(12)=PYRITE
      XEND(13)=UPERTH
      XEND(14)=UO2TON
      XEND(15)=RAINTH
      XEND(16)=THDISS
      XEND(17)=RAINGP
      XEND(18)=CONCRA
      XEND(19)=THFRAC
      XEND(20)=XK1
      XEND(21)=XK3
      XEND(22)=BETA
      XEND(23)=BF
      XEND(24)=MPCH2O
      XEND(25)=EMANAT
      XEND(26)=FPREC
      XEND(27)=FWIND1
      XEND(28)=FWIND2
      XEND(29)=HSRC
      XEND(30)=UWIND
      XEND(31)=VDISP
      XEND(32)=DEPVEL
      XEND(33)=FDEPSF(1)
      XEND(34)=FDEPSF(2)
      XEND(35)=FDEPSF(3)
      XEND(36)=FDEPSF(4)
      XEND(37)=FDEPED(1)
      XEND(38)=FDEPED(2)
      XEND(39)=FDEPED(3)
      XEND(40)=FDEPED(4)
      XEND(41)=DUREXP(1)
      XEND(42)=DUREXP(2)
      XEND(43)=DUREXP(3)
      XEND(44)=DUREXP(4)
      XEND(45)=VEGLAM(1)
      XEND(46)=VEGLAM(2)
      XEND(47)=VEGLAM(3)
      XEND(48)=VEGLAM(4)
      XEND(49)=YDEN(1)
      XEND(50)=YDEN(2)
      XEND(51)=YDEN(3)
      XEND(52)=YDEN(4)
      XEND(53)=FBI
      XEND(54)=FAUT(1)
```

Table 5.3 continued

```
      XEND(55)=FAUT(2)
      XEND(56)=FACTA
      XEND(57)=FACTB
      XEND(58)=FLOCA
      XEND(59)=FMEAT
      XEND(60)=FISHIN
      XEND(61)=OF
      XEND(62)=F
       RETURN
       END
C
       SUBROUTINE TIMOUT
$INCLUDE:'B:STUFF.INC'
      XTIME1(NCOUNT)=VGDTOT
      XTIME2(NCOUNT)=XMEATD
      XTIME3(NCOUNT)=WATDOS
      XTIME4(NCOUNT)=FISHDS
      XTIME5(NCOUNT)=XINDOS
      XTIME6(NCOUNT)=RNDOS
      XTIME7(NCOUNT)=TOTDOS
      XTIME8(NCOUNT)=PH(5,2)
C       WRITE(7,10)XI1,T,VGDTOT,XMEATD,WATDOS,FISHDS,XINDOS,RNDOS,TOTDOS
C  10    FORMAT(1X,F8.4,8E13.6)
        RETURN
       END
C
C
C       SUBROUTINE SUBDAT
C$INCLUDE:'B:STUFF.INC'
C       READ(3,900) RHOTAL,NPRCP,DUNSAT,DMIX,WVOID,TVOID,WCONC(1,4,2),
C                  CONC(7,1,2),CONC(7,2,2),DEPTH(2),WCONC(4,4,2),
C     ;            WCONC(9,4,2),WCONC(7,4,2),WCONC(8,4,2),PH(4,2),
C     ;            SULPRD,O2DIFF,KPRIME,FACTCO,DISTCO,RAINGP,CONCRA,
C     ;            THFRAC
C 900   FORMAT(1X,23(E13.6,1X))
C       RETURN
C       END
```

Table 5.4

$INCLUDE File That Contain COMMON Blocks Used In PAC and Other Subroutines

```
C     FILE 'STUFF.INC' CONTAINS INTERMEDIATE RESULTS,INPUT AND
C     OUTPUT FOR TRANSFER FROM 'SOURCE' TO OTHER PROGRAM MODULES.
      REAL*4 KPRIME,KSUBSO,NPRCP,KSUBDR,KRATE,MPCH2O
      REAL*4 PHIN(5),FRACRA(2)
      COMMON/CTIME/IYR,T,NCOUNT,NDT,XI1
C PRINTER CODES
      COMMON/INTERM/ICODS,ICODL,ICODPH,ICODF,ICODR,ICODD,ICODA,ICODV1
     &,ICODV2,ICODO
C SOURCE-INTERMEDIATE QUANTITIES
      COMMON/INTERM/CONC(10,4,2),WCONC(10,5,2),ACALC,KRATE,RAFRNI,
     ;       TFINAL,DELTAH(2),CONCRG,FRACR1,CONCG,PH(5,2),DEPTH(2)
     ;       ,SFLUX
C RADON-INTERMEDIATE QUANTITIES
      COMMON/INTERM/EBKDFC,DIFLEN,RNFLUX
C LAKE-INTERMEDIATE QUANTITIES
      COMMON/INTERM/ A13,A22,A21,RAWAT,RASED,VLAKE,XK2
C DUST-INTERMEDIATE QUANTITIES
      COMMON/INTERM/EMRATE,DTEMP
C AIRDIS-INTERMEDIATE QUANTITIES
      COMMON/INTERM/CAIR,TEMEXP,ATEMP,CAIRN
C VEGDEP-INTERMEDIATE QUANTITIES
      COMMON/INTERM/CVEG(4),DUMVEG(4)
C VEGEAT- INTERMEDIATE QUANTITIES
      COMMON/INTERM/CMEAT,DUMEAT
C FISH-INTERMEDIATE QUANTITIES
      COMMON/INTERM/CFISH
C  DOSEIN-INTERMEDIATE QUANTITIES
      COMMON/INTERM/DUMVGA,DUMVGB,DUMET,DUMWAT,DMFISH,DUMINH,
     &VGDOS3,VGDOS4,VGDTOT,XMEATD,WATDOS,FISHDS,XINDOS,RNDOS
     &,TOTDOS
C  TIME  PARAMETERS AND PRINT CONTROL CODES
      COMMON/INPT/FWIND1,FWIND2,FPREC,UWIND,RYEARS,DT,OTIME
C  PRINTER INPUTS
      COMMON/INPT/CODS,CODL,CODPH,CODF,CODR,CODD,CODA,CODV1,CODV2,CODO
     &,CODU
C SOURCE-INPUTS
      COMMON/INPT/SULPRD,RHOTAL,O2CONC,O2DIFF,KPRIME,KSUBSO,NPRCP,
     ;       FACTCO,DISTCO,DUNSAT,DMIX,KSUBDR,RAINGP,TVOID,
     ;       CONCRA,FRACGP,PYRITE,UPERTH,UO2TON,RAINTH,
     ;       THDISS,THFRAC
C  LAKE-INPUTS
      COMMON/INPT/XK1,XK3,BETA,ALAKE,DLAKE,AWAT,ATAIL
C  LAKEPH-INPUTS
      COMMON/INPT/ACTCO
C  FISH-INPUTS
      COMMON/INPT/BF
C  RADON-INPUTS
      COMMON/INPT/MPCH2O,RNLAM,EMANAT
C  DUST-INPUTS
      COMMON/INPT/SILT,AEXPT
C  AIRDIS-INPUTS
      COMMON/INPT/HSRC,XDIST,VDISP
C  VEGDEP-INPUTS
      COMMON/INPT/FDEPSF(4),FDEPED(4),DUREXP(4),DEPVEL,VEGLAM(4)
     &,YDEN(4)
C  VEGEAT-INPUTS
      COMMON/INPT/FBI,FAUT(2),FIR
C  DOSEIN-INPUTS
      COMMON/INPT/DCFING,VEGINA,VEGINB,FACTA,FACTB,FLOCA,
     &XMEATI,FMEAT,WATIN,FISHIN,FFISH,DCFINH,XINHIN,OF,F
C  OUTPUT ARRAY
      COMMON/RESULT/XTIME1(11),XTIME2(11),XTIME3(11),XTIME4(11)
     &,XTIME5(11),XTIME6(11),XTIME7(11),XTIME8(11),XEND(62)
```

Table 5.5

Dictionary for PAC, ENDOUT and TIMOUT

Variable	Description
I1	Trial Number
RYEARS	Simulation Length
DT	Time increment Δt
CODS	Detailed print code for source subroutine
CODPH	Detailed print code for lake pH subroutine
CODL	Detailed print code for lake subroutine
CODF	Detailed print code for fish concentration subroutine
CODR	Detailed print code for radon subroutine
CODD	Detailed print code for dust subroutine
CODA	Detailed print code for air dispersion subroutine
CODV1	Detailed print code for annual consumption of vegetation subroutine
CODO	Detailed print code for dose subroutine
COUNT	Counter to obtain time dependent output
IYR	Time loop counter (a)
OTIME	Time interval for saving of time-dependent output
XEND(62)	Array to store input parameter to be saved (C2 storage location)
XTIME (11)	Array to store 11 vegetation dose values
XTIME2 (11)	Array to store 11 meat consumption dose values
XTIME3 (11)	Array to store 11 water consumption dose values
XTIME4 (11)	Array to store 11 fish consumption dose values
XTIME5 (11)	Array to store 11 dust inhalation dose values
XTIME6 (11)	Array to store 11 radon inhalation dose values
XTIME7 (11)	Array to store 11 total dose values
XTIME8 (11)	Array to store 11 lake pH values

See model writeups in Appendices A to F for description of parameters in subroutine ENDOUT.

To change the parameters stored in the output vector, the procedure outlined in Figure 5.2 should be used. As indicated on the figure, the first question to be addressed is whether the desired parameter is accessible in a COMMON block. Parameters currently accessible via common blocks are indicated in Table 5.4. Most calculated parameters which might be of interest are present in the COMMON/INTERM/ block. Parameter definitions and descriptions as well as model documentation are presented (along with their FORTRAN variable names) in Appendices A to F. If the parameter is noted in a common block, it is available to PAC and only a simple programming change is required.

To change either input or calculated non-time-dependent parameters for transfer to RANSIM the following procedure should be used. In ENDOUT (the output subroutine for non-time-dependent parameters), a position in the vector XEND must be set equal to the variable desired for output (see SUBROUTINE ENDOUT in Table 5.3). XEND must then be respecified in correctly-dimensioned form in the common block RESULT in the $INCLUDE file B:STUFF.INC. In this location, it references the vector XOUT in RANSIM via the same common block.

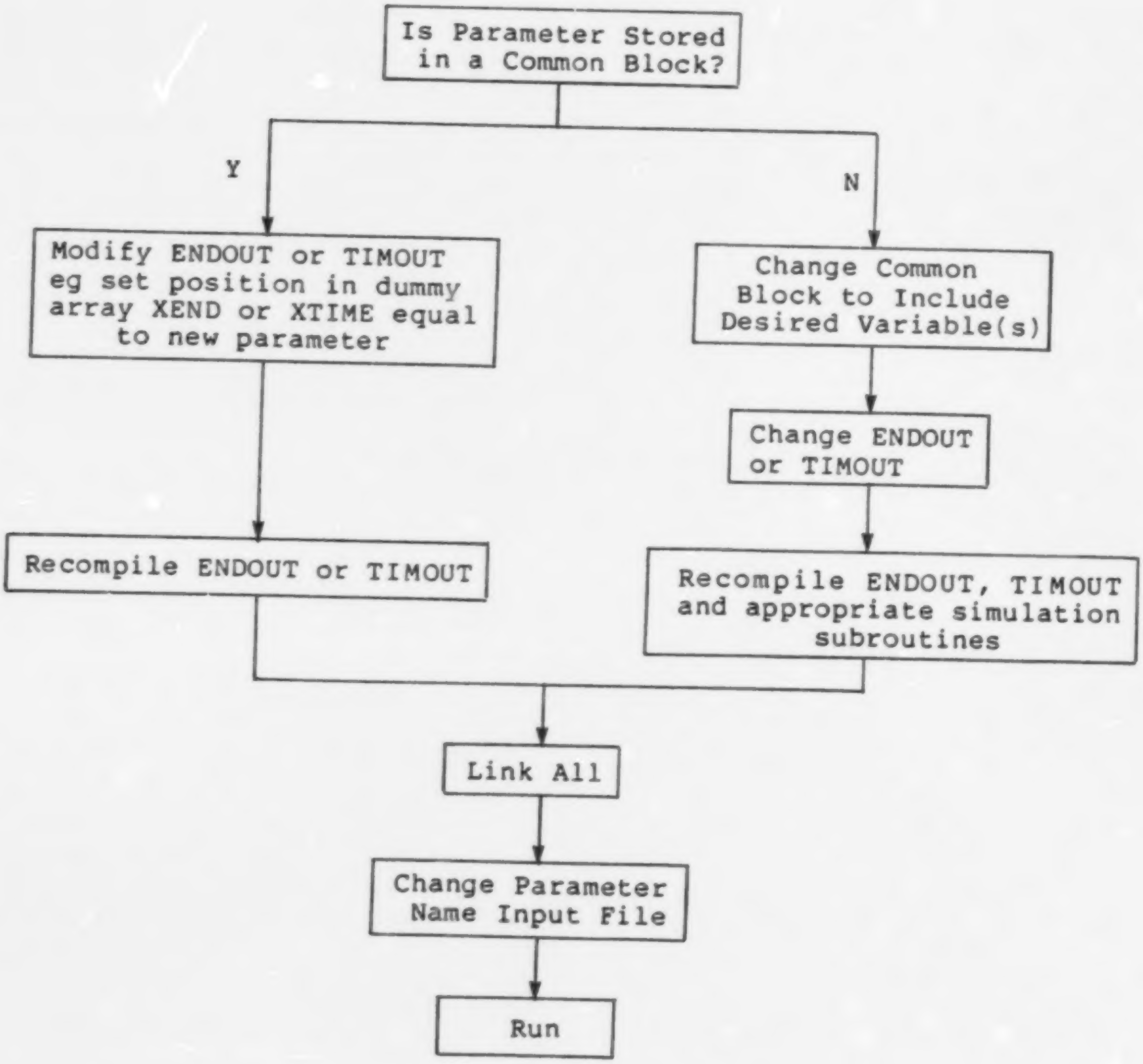
The procedure for modifying input or calculated time-varying parameters is somewhat different. In SUBROUTINE TIMOUT, a series of vectors called XTIME have OTIME positions to be filled. For example, to fill a position in the array XTIME1 with a desired parameter requires a statement such as:

```
XTIME1 (NCOUNT) = variable name
```

It is also necessary to add XTIME1 to the common block RESULT in the $INCLUDE file B:STUFF.INC. Note that there are a maximum of 150 storage locations in RESULT.

FIGURE 5·2

PROCEDURE USED TO CHANGE PARAMETER WRITTEN TO OUTPUT FILE


Is Parameter Stored in a Common Block?
Y
N
Modify ENDOUT or TIMOUT eg set position in dummy array XEND or XTIME equal to new parameter
Change Common Block to Include Desired Variable(s)
Change ENDOUT or TIMOUT
Recompile ENDOUT or TIMOUT
Recompile ENDOUT, TIMOUT and appropriate simulation subroutines
Link All
Change Parameter Name Input File
Run

Once the necessary programming changes have been made, the affected subroutine(s) must be recompiled to create a new object module. This revised object module can then be linked with the previously created object modules to produce an executable program. Prior to running the revised program, the file containing names and descriptions of output parameters should be modified to reflect the programming changes made.

If a parameter desired for output is not listed in a common block, the procedure is somewhat more involved. The file STUFF.INC which contains all of the common blocks, must be edited and the desired parameter added to one of the existing common blocks. The programming changes that follow are then the same as the changes detailed above for non-time-dependent and time-dependent parameters except for the following. Since STUFF.INC appears in most of the program modules, it is advisable, after changing STUFF.INC to recompile all source programs. Once all necessary revised object modules have been created they can be linked to produce an executable program. Again, the parameter name file should be revised to reflect the programming changes made.

6.0 QUALITY ASSURANCE

Quality Assurance (QA) is an integral part of computer software development, particularly for large models such as UTAP. Consistent, logical procedures must be employed at all stages of program development to ensure program integrity. Software QA is not a rote, mechanical process but rather a creative process that has unique features for each different application. An overview of how QA concepts and techniques were applied during the development of UTAP is provided below. Some specific QA tests of the random number generator and sampling routines are presented in Section 3.6 of Volume 1. A more detailed review of current concepts in software QA is presented in the Phase I study (SENES 1984).

Like many other large models, UTAP consists of a number of components that have been developed (and coded) by various contributors. Therefore methods were needed to ensure the integrity of the component (simulation) models as well as the system model which is composed of the control modules (RANSIM and PAC) and the component models.

Component model development and QA involve the specification of goals (requirements), formulation of equations and definition of input values for each component model. It is particularly important to define the inputs probabilistically at an early stage so that the models are developed to properly incorporate the full range of input values. During the Phase II studies, the component model descriptions and formulations were subjected to extensive peer review prior to implementation. Large submodels were implemented and tested individually, using RANSIM and an appropriate PAC module. This facilitated eventual integration into the overall system model.

Detailed printout statements (able to be executed by a simple print code) were included at all critical points in the component models. The print codes indicate all inputs as well as calculated quantities. These diagnostic tools facilitated error identification and correction during the development of UTAP.

System QA deals with the integration of the control modules and component models. Component models were integrated into the probabilistic system model in a logical sequence, corresponding to the order in which calculations were performed. Once integrated, all aspects of the system model were fully tested. Individuals responsible for the development of particular component models checked that results generated by their models were reasonable and correct.

In this project, the SOURCE model for the tailings pile was developed, tested and implemented first. Since the model was time-consuming to run, an output file from SOURCE was generated and used as a surrogate in the development of the other component models. Once the "downstream" models were developed and subjected to QA all the component models were folded together. The system QA program detailed below was essentially a feedback loop between the system modeller and the component model formulators.

The system model was first run deterministically, with all parameters held constant at their nominal values. The individual(s) responsible for the development of each model were given a copy of the input, the code, and the detailed printout to check. Mistakes were corrected and the system model was run deterministically for two trials for two hundred years. The inputs, code, and output were again checked by the formulators. In this case the time performance of the program was checked. As well, the two deterministic trials were checked to ensure that they yielded the same results. This provided a preliminary check on the ability of the system model to perform successive trials.

As the QA program continued, the system model was run probabilistically for two trials. Inputs for each trial were checked to ensure that they were being drawn from the correct distributions. A series of small runs of 20 trials were done and several significant output parameters were examined for agreement with the state of knowlege about the physical system. Discrepancies between the two were resolved. This process continued until all parties were satisfied that the models were functioning properly.

As a further QA measure, the system model was again run deterministically; this time using extreme high or low values for all input parameters, to examine the extreme response of the system. Again, any problems were corrected. Finally, the system model was run for many trials (500) and many years (1000). Since the model utilizes probabilistic inputs, running 500 trials allows an assessment to be made of the ability of the model to produce correct results from many combinations of the input parameters. The 1,000-year simulation period ensures that the model will properly simulate this period of time. Although there is a high degree of confidence that the system model will properly respond to most combinations of input parameters, it still remains possible that infeasible combinations of parameters leading to erroneous results or system failure can still occur. QA does not excuse users of the code from examining the inputs and the results in a critical manner.

7.0 REFERENCES

Reilly, P.M. and Johri, H.P., 1969. "Decision Making Through Opinion Analysis." Chemical Engineering, April.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada - Phaes I." Research Report prepared for the National Uranium Tailings Program, CANMET, EMR, under Supply and Services Canada Contract No. OSQ83-00236.

APPENDIX A

THE CALCULATION OF THE WATER SOURCE TERM
(SUBROUTINE SOURCE)

APPENDIX A
TABLE OF CONTENTS

Page

1.0 INTRODUCTION A-1

2.0 INITIAL MASS BALANCE CONSIDERATIONS A-2
- 2.1 Initial Gypsum Inventory A-2
- 2.2 Initial Radium Inventory A-3
- 2.3 Initial Thorium Inventory A-3

3.0 DETERMINATION OF DISSOLVED RADIUM AS A FUNCTION OF DEPTH AND TIME (UNSATURATED ZONE) A-5
- 3.1 Determination of the Radium Distribution Coefficient K A-6
- 3.2 Determination of Dissolved Gypsum Concentration A-9
- 3.3 Determination of Sulphates Derived from Pyrite Oxidation A-11
- 3.4 Determination of Mole Fraction of Radium in Reprecipitated Gypsum A-14
  - 3.4.1 Depth of Depleted Zone and Time to Total Depletion A-15
  - 3.4.2 Reprecipitated Gypsum Content (Enriched Layer) A-17
  - 3.4.3 Radium Content of Reprecipitated Gypsum A-18

4.0 POREWATER pH AS A FUNCTION OF TIME A-20

Table of Contents (continued)

Page

5.0 DETERMINATION OF THORIUM CONCENTRATION AND RADIUM INGROWTH FROM THORIUM AS FUNCTIONS OF TIME A-21
  5.1 Thorium Concentration A-21
    5.1.1 In Solid Phase A-21
    5.1.2 In Liquid Phase A-21
  5.2 Radium Ingrowth from Thorium A-22
    5.2.1 In Liquid Phase A-22
    5.2.2 In Solid Phase A-23

6.0 DETERMINATION OF RADIUM ASSOCIATED WITH THE SOLID PHASE AS A FUNCTION OF TIME AND DEPTH A-24

7.0 EFFECTS OF THE MIXING ZONE A-25
  7.1 Ionic Concentrations and pH Calculation A-26
    7.1.1 Calcium, Sulphate and Iron Concentrations A-26
    7.1.2 pH Calculation A-29
  7.2 Radionuclide Concentrations in the Mixing Zone (Liquid Phase) A-29
    7.2.1 Radium Concentration A-29
    7.2.2 Thorium Concentration A-30

8.0 MODEL REVISIONS FOR ALTERNATIVE SCENARIOS A-31
  8.1 Scenario 2: Pyrite Reduction A-31
  8.2 Scenario 3: Pyrite Reduction and Limestone Addition A-32
  8.3 Scenario 4: Pyrite and Radionuclide Reduction and Limestone Addition A-38

REFERENCES A-39

APPENDIX A

THE CALCULATION OF THE WATER SOURCE TERM (SUBROUTINE SOURCE)

1.0 INTRODUCTION

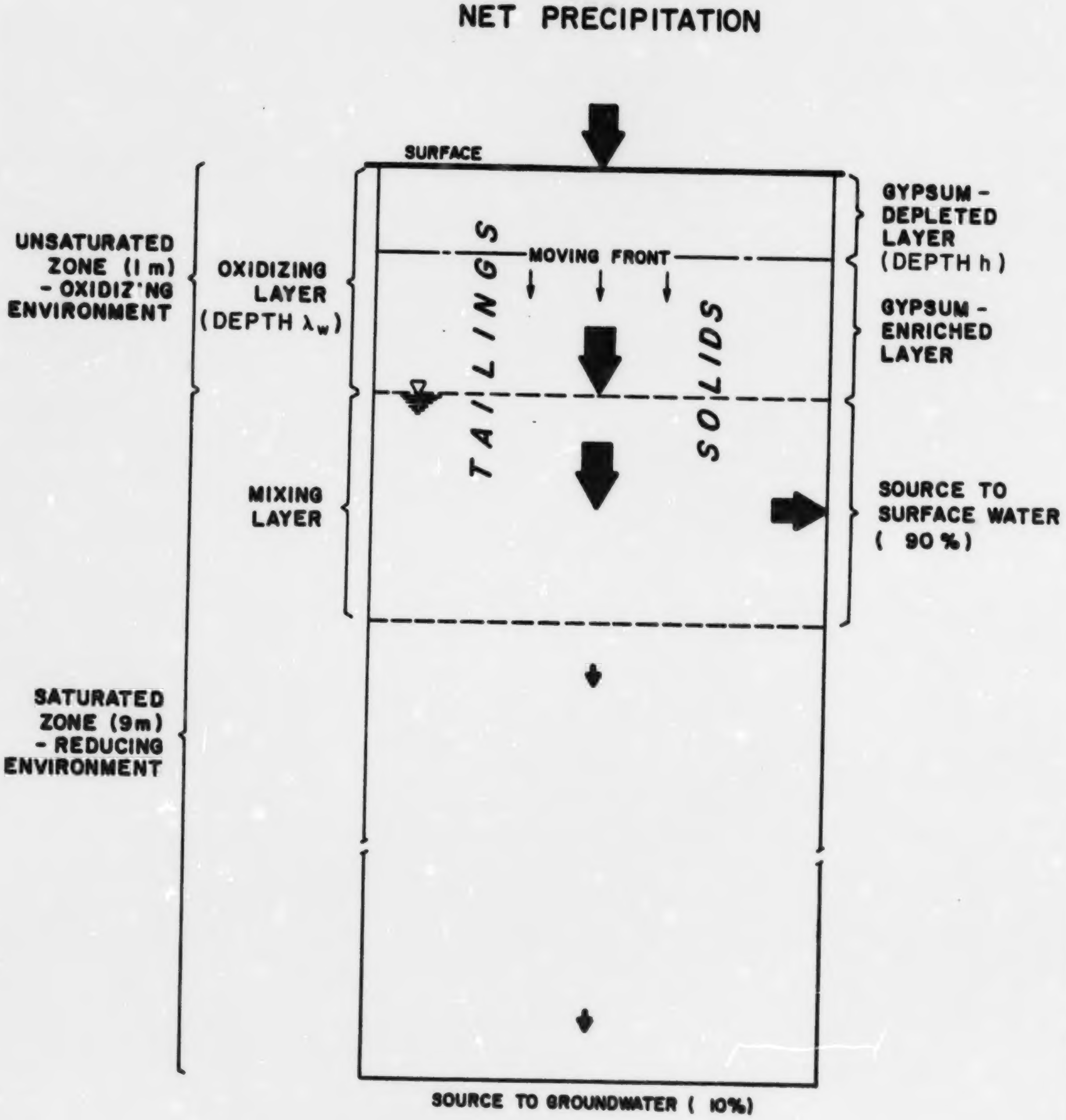
In Phase II, the tailings source is modelled as a layered system, as shown in Figure A.1. The features shown on the figure are explained in more detail in subsequent sections. Briefly, the reference tailings site described in the Phase I report is further explored in Phase II. The tailings consist of an upper 1 metre deep unsaturated (oxidizing) zone, and below this a 9 metre deep saturated zone.

As time progresses, the gypsum content of the reference tailings material is gradually depleted. Gypsum removal is hypothesized to proceed as a front moving downward through the tailings. Consequently, the unsaturated zone can be further subdivided into an upper layer, depleted of gypsum and a lower layer, enriched in gypsum. The relative height of these two layers changes with time.

The upper 1 m of the saturated zone has been specified for modelling purposes as a 'mixing layer' in which adsorption of several species may occur. This feature is necessary to account for the natural retardation processes which occur in large tailings areas. The time varying behaviour of a number of chemical species is modelled, in both solid and liquid phases, at several depths in the tailings. Species modelled include sulphate and gypsum, pH, radium and thorium concentrations. A summary listing of the parameters used in the water source term model and their assigned probability distributions is provided in Table A.1.

FIGURE A·1

CONCEPT ADOPTED FOR MODELLING THE TAILINGS SOURCE TERM


NET PRECIPITATION
SURFACE
TAILINGS
SOLIDS
MOVING FRONT
UNSATURATED ZONE (1 m) - OXIDIZ'NG ENVIRONMENT
OXIDIZING LAYER (DEPTH λ_w)
MIXING LAYER
SATURATED ZONE (9m) - REDUCING ENVIRONMENT
GYPSUM - DEPLETED LAYER (DEPTH h)
GYPSUM - ENRICHED LAYER
SOURCE TO SURFACE WATER (90 %)
SOURCE TO GROUNDWATER (10%)

2.0 INITIAL MASS BALANCE CONSIDERATIONS

Characterization of the physical and chemical behaviour of a tailings area requires specification of the initial inventories of many chemical species. To achieve consistency in the mass balance calculations, several fundamental parameters were defined as follows:

ρ_T = tailings bulk density (g.m^{-3})
Ra_s^o = initial total activity of Ra-226 in the tailings (Bq.g^{-1})
$X_{Ra/gyp}$ = fraction of total Ra-226 which is associated with gypsum
X_{gyp} = mass fraction of gypsum in tailings
$[FeS_2]^o$ = initial concentration of pyrite in tailings (mol.m^{-3})
$X_{U/Th}$ = mass ratio of UO_2 to ThO_2 in ore
X_{UO_2} = initial (mass) fraction of UO_2 per tonne of ore (kg UO_2.t^{-1} ore)
$X_{Ra/Th}$ = mol fraction of Ra-226 ingrowth to solid Th-230 matrix which is insoluble.
X_{Th} = fraction of total Th which is assumed to be immobile (ie. fraction of Th assumed to be in crystalline rather than amorphous form)

These parameters were used as the basis for calculating the initial inventories described below.

2.1 Initial Gypsum Inventory

$$g_o = \rho_T \; X_{gyp} \left[\frac{1 \text{ mol } CaSO_4}{136 \text{ g}}\right] \qquad (2.1)$$

where

g_o = initial solid phase gypsum content in bulk tailings (mol.m^{-3})

2.2 Initial Radium Inventory

$$[Ra_s]^o = Ra_s^o \; \rho_T \left(\frac{27 pCi}{Bq}\right)\left(\frac{1 \; g \; Ra}{10^{12} \; pCi}\right)\left(\frac{1 \; mol \; Ra}{226 \; g \; Ra}\right) \qquad (2.2)$$

$$= 1.196 \times 10^{-13} \quad Ra_s^o \; \rho_T$$

where

$[Ra_s]^o$ = total initial concentration of radium in tailings solid phase ($mol.m^{-3}$)

Then:

$[Ra_s]^o_{gyp} = \quad x_{Ra/gyp} \quad [Ra_s]^o$

= initial concentration of Ra associated with gypsum ($mol.m^{-3}$ tailings);

and

$[Ra_s]^o_m = (1 - x_{Ra/gyp}) \quad [Ra_s]^o$

= initial concentration of Ra associated with host tailings material assumed not to be leachable ($mol.m^{-3}$ tailings);

and

$\bar{x}^o_{Ra/gyp} = [Ra_s]^o_{gyp}/g_o$

= initial mol fraction of Ra-226 in gypsum

2.3 Initial Thorium Inventory

The initial activity of Th-230 is determined by assuming secular equilibrium between Ra-226 and its parent Th-230. Hence, the initial activity of Th-230 must equal the initial activity of Ra-226. Since activity can be expressed as the product of a species' concentration and its decay constant, one may write:

$$\lambda_{Ra-226}[Ra_s]^o = \lambda_{Th-230}[Th\text{-}230_s]^o$$

where the decay constants are respectively:

$$\lambda_{Ra-226} = \frac{\ln 2}{t_{1/2}Ra\text{-}226} = \frac{\ln 2}{1600\ yr} = 4.33 \times 10^{-4}\ yr^{-1}$$

$$\lambda_{Th-230} = \frac{\ln 2}{t_{1/2}Th\text{-}230} = \frac{\ln 2}{80{,}000\ yr} = 8.66 \times 10^{-6}\ yr^{-1}$$

Thus, the initial solid phase concentration of Th-230 can be calculated from:

$$[Th\text{-}230_s]^o = \frac{\lambda_{Ra-226}}{\lambda_{Th-230}} [Ra_s]^o \qquad (2.3)$$

Th-230 is considered to be present, along with the much longer lived Th-232 ($t_{1/2} = 1.41 \times 10^{10}$ yr) as $Th(OH)_4$, which is formed during the milling process. Thus any Ra-226 which grows in from Th-230 decay will be present in the $Th(OH)_4$ matrix. In order to estimate the [Ra] in the $Th(OH)_4$ matrix, it is necessary to know the mol fraction of Th-230 in the $Th(OH)_4$ matrix.

The initial concentration of Th-232 is calculated as:

$$[Th\text{-}232_s]^o = ({}^{X}UO_2/{}^{X}U/Th)\ (\rho_T/10^6)\ (mol\ ThO_2/264\times10^{-3}kg) \qquad (2.4)$$

where

$[Th\text{-}232_s]^o$ = the initial solid phase Th-232 content of the tailings ($mol.m^{-3}$)

The total solid phase concentration of thorium is given by:

$$[Th_s]^o = [Th\text{-}230_s]^o + [Th\text{-}232_s]^o \qquad (2.5)$$

The mole fraction of Th-230 to Th-232 is then:

$$X_{Th/Th} = \frac{[\text{Th-230}_s]^o}{[\text{Th-232}_s]^o} \qquad (2.6)$$

3.0 DETERMINATION OF DISSOLVED RADIUM AS A FUNCTION OF DEPTH AND TIME (UNSATURATED ZONE)

One of the main sources of dissolved radium is from the dissolution of gypsum. Gypsum forms a significant component of uranium mill tailings as a result of lime additions for pH adjustment in the mill. The calcium and radium sulphate compounds are assumed to form a solid co-precipitate (non-ideal solid solution) $Ra_xCa_{1-x}SO_4$ (solid). As well, radium may co-precipitate with other sulphates (eg. $BaSO_4$, $PbSO_4$). However, the distribution of radium between these compounds is not well known, hence gypsum is used here as a surrogate for the many sulphate co-precipitates which may be present. This solid solution is in equilibrium with a liquid phase; the tailings porewater. The co-precipitation of the dissolved Ra-226 with a solid phase other than gypsum ($BaSO_4$ or $PbSO_4$, for example) would require a relatively minor re-definition of the equilibria controlling Ra-226 transport, since the common ion (i.e. sulphate) continues to depend on pyrite oxidation and gypsum dissolution mechanisms as modelled herein. Using gypsum as the co-precipitate a relationship can then be developed as such:

$$y = Kx \qquad (3.1)$$

where

y = mole fraction of radium in the liquid phase

$$= \left.\frac{[Ra^{2+}]}{[Ra^{2+}]+[Ca^{2+}]}\right|_{\text{liquid}} \quad \left.\frac{[Ra^{2+}]}{[Ca^{2+}]}\right|_{\text{liquid}}$$

x = mole fraction of radium in the solid phase

$$= \left.\frac{[Ra^{2+}]}{[Ra^{2+}]+[Ca^{2+}]}\right|_{solid} \qquad \left.\frac{[Ra^{2+}]}{[Ca^{2+}]}\right|_{solid}$$

K = distribution coefficient for radium between solid and liquid phases

Multiplying equation (3.1) by the concentration of gypsum in solution ($[gyp_{(aq)}]=[CaSO_4]=[Ca^{2+}(aq)]$) the radium concentration in solution can be expressed as:

$$[Ra^{2+}(aq)] = Kx\ [Ca^{2+}(aq)] \tag{3.2}$$

or

$$[Ra^{2+}(aq)] = Kx\ [gyp\ (aq)]$$

Thus, to determine $[Ra^{2+}(aq)]$, (i.e. $[Ra^{2+}]$ in porewater) the values of K, x and [gyp(aq)] must be determined.

It should be noted that equation (3.2) is only valid as long as gypsum remains in the unsaturated zone. As will be explained in subsequent sections, gypsum in the unsaturated zone dissolves and becomes totally depleted over time. Once this depletion occurs, the source term chemistry is simplified considerably, and an alternative model must be used. The alternative model, for use after gypsum depletion has occurred, is presented in section 3.3.3.

3.1 Determination of the Radium Distribution Coefficient K

If $CaSO_4$ and $RaSO_4$ form a regular, but non-ideal solid solution, (Lippmann 1977, 1980; Gresens 1981), then:

$$y = \frac{K_{s,Ra}\ \exp[\alpha(1-x)^2]x}{K_{s,Ra}\ \exp[\alpha(1-x)^2]x+K_{s,Ca}\ \exp[\alpha x^2](1-x)} \tag{3.3}$$

where

$$K_{s,Ra} = \frac{K^{o}_{s,RaSO_4}}{(f_{Ra^{2+}})(f_{SO_4^{2-}})} \qquad K_{s,Ca} = \frac{K^{o}_{s,CaSO_4}}{(f_{Ca^{2+}})(f_{SO_4^{2-}})}$$

$K^{o}_{s,RaSO_4}$ = solubility product (zero ionic strength) of $RaSO_4$ $(mol^2.m^{-6})$

$K^{o}_{s,CaSO_4}$ = solubility product (zero ionic strength) of $CaSO_4$ $(mol^2.m^{-6})$

$f_{Ra^{2+}}, f_{Ca^{2+}}, f_{SO_4^{2-}}$ = activity coefficients (liquid phase)

= coefficient of non-ideality

Equation (3.3) can be simplified for x<<1 to yield:

$$y = \frac{1}{1 + \left[\dfrac{K^{o}_{s,CaSO_4} \quad f_{Ra^{2+}}}{K^{o}_{s,RaSO_4} \quad f_{Ca^{2+}} \quad \exp(\alpha) x}\right]} \qquad (3.4a)$$

If the following values are assigned to the parameters in equation (3.4a):

$K^{o}_{s,RaSO_4} = 10^{-10.37} = 4.26 \times 10^{-11}$ $(mol^2.L^{-2})$ (Bjerrum et al. 1958)

$K^{o}_{s,CaSO_4} = 2.5 \times 10^{-5}$ $(mol^2.L^{-2})$

then

$$\frac{K^{o}_{s,\ CaSO_4}}{K^{o}_{s,RaSO_4}} = 5.85 \times 10^5$$

$$\frac{f_{Ca^{2+}}}{f_{Ra^{2+}}} = 1.0$$

and: $\alpha = 9.0$ to 11.0

It is apparent that the right hand side of the denominator of equation (3.4a) is much greater than 1, hence equation (3.4a) may be reduced to:

$$y = \left(\frac{K^{o}_{s,RaSO_4}}{K^{o}_{s,CaSO_4}}\right)\left(\frac{f_{Ca^{2+}}}{f_{Ra^{2+}}}\right) \exp(\alpha)x \qquad (3.4b)$$

If we let

$$\left(\frac{K^{o}_{s,RaSO_4}}{K^{o}_{s,CaSO_4}}\right)\left(\frac{f_{Ca^{2+}}}{f_{Ra^{2+}}}\right) \exp(\alpha) = K$$

then we have once again the expression:

$$y = Kx \qquad (3.5a)$$

Substitution of the values given above for the parameters in equation (3.5a) yields:

$K = 0.01$ to 0.1

and

$y = 0.01\,x$ to $0.1\,x$ (3.5b)

3.2 Determination of Dissolved Gypsum Concentration

The solubility of gypsum, and hence the concentration of $[Ca^{2+}]$ is dependent on the solubility product as follows:

$$\frac{K^{o}_{s,CaSO_4}}{\left(f_{Ca^{2+}}\right)\left(f_{SO_4^{2-}}\right)} = \left[SO_4^{2-}\right]\left[Ca^{2+}\right] \qquad (3.6)$$

where

$K^{o}_{s,CaSO_4}$ = solubility equilibrium constant (25.0 $mol^2.m^{-6}$)

$f_{Ca^{2+}}$ = activity coefficient (calcium ion)

$f_{SO_4^{2-}}$ = activity coefficient (sulphate ion)

The activity coefficients, assumed to be essentially equal (i.e. f) depend on a number of factors (e.g. ionic strength of solution, temperature, etc.) and are assigned an appropriate distribution in the probabilistic assessment.

It may be assumed that the limestone ($CaCO_3$) content of pyritic tailings is rapidly depleted due to acid generation resulting from pyrite oxidation above the water table. Thus, the dissolution of gypsum is the major source of calcium during the initial years. On the other hand, pyrite oxidation of tailings contributes significantly to the sulphate concentration in the porewater. That is:

$$[Ca^{2+}] = [Ca^{2+}]_g \qquad (3.7)$$

$$[SO_4{}^{2-}] = [SO_4{}^{2-}]_g + [SO_4{}^{2-}]_p$$

where

$[Ca^{2+}]_g$ = dissolved calcium derived from gypsum ($mol.m^{-3}$)
$[SO_4^{2-}]_g$ = dissolved sulphate derived from gypsum ($mol.m^{-3}$)
$[SO_4^{2-}]_p$ = dissolved sulphate derived from pyrite oxidation ($mol.m^{-3}$)

In the relationship expressed above:

$$[Ca^{2+}]_g = [SO_4^{2-}]_g$$

hence, substituting into the equation given previously for the solubility product (equation (3.6)) the following expression is obtained:

$$\frac{K^{o}_{s,CaSO_4}}{f^2} = ([SO_4{}^{2-}]_p + [SO_4{}^{2-}]_g)\ [Ca^{2+}]_g \qquad (3.8)$$

or

$$\frac{K^{o}_{s,CaSO_4}}{f^2} = [SO_4{}^{2-}]_p\ [Ca^{2+}]_g + [Ca^{2+}]_g^2$$

Solving equation (3.8) for $[Ca^{2+}]$ (or $[Ca^{2+}]_g$) gives:

$$[Ca^{2+}] = \sqrt{\frac{K^{o}_{s,CaSO_4}}{f^2} + .25\ [SO_4^{2-}]_p^2} - 0.5\ [SO_4^{2-}]_p \qquad (3.9)$$

Thus, determination of the $[Ca^{2+}]$ depends on the concentration of sulphate due to pyrite oxidation.

3.3 Determination of Sulphates Derived from Pyrite Oxidation

The stoichiometry of the bacteria-assisted pyrite oxidative process is given by:

$$2\ FeS_2 + 7.5\ O_2 + 7\ H_2O \rightarrow 2\ Fe(OH)_3 + 8\ H^+ + 4\ SO_4^{2-} \qquad (3.10)$$

From the stoichiometric relationship it is seen that the rate of sulphate formation is related to the rate of pyrite oxidation by the expression:

$$\frac{1}{2}\frac{d[SO_4^{2-}]}{dt} = -\frac{d[FeS_2]}{dt} \qquad (3.11)$$

In the above equation, the rate of pyrite oxidation can be expressed as:

$$-\frac{d[FeS_2]}{dt} = k\ [FeS_2]$$

and the pyrite concentration in the solid phase can be calculated from:

$$[FeS_2] = [FeS_2]^{\circ}\ \exp(-kt)$$

Then on substitution in equation (3.11) one gets:

$$-\frac{d[FeS_2]}{dt} = k\ [FeS_2]^{\circ}\ \exp(-kt) = \frac{1}{2}\frac{d[SO_4^{2-}]}{dt}$$

At time t=0 the equation reduces to:

$$k = \left.\frac{\frac{1}{2}\frac{d[SO_4^{2-}]}{dt}}{[FeS_2]^{\circ}}\right|_{t=0}$$

In the latest expression, the rate of sulphate formation can be restated as:

$$\left.\frac{d[SO_4^{2-}]}{dt}\right|_{t=0} = \frac{J_s}{\lambda_w}$$

where

λ_w = depth of unsaturated zone (m)
J_s = net sulphate flux ($mol.m^{-2}.yr^{-1}$)

An expression for J_s has been previously derived (SENES 1984). It can be shown that the steady state sulphate flux (J_s) derived from the oxygen flux into the tailings is given by:

$$J_s = \left[\frac{8}{7.5} \quad \bar{R}_s \quad \rho_T \quad [O_2] \quad D_{O_2} \left(1 - K' \ln\left(\frac{1+K'}{K'}\right)\right)\right]^{1/2} \qquad (3.12)$$

where

$\bar{R}_s$ = sulphate production rate ($mol.g^{-1}.yr^{-1}$)
$[O_2]$ = oxygen concentration at surface ($mol.m^{-3}$)
D_{O_2} = oxygen diffusivity through tailings ($m^2.yr^{-1}$)
K' = $K_{O_2}/[O_2]$ = ratio of O_2 half saturation constant (K_{O_2}) to $[O_2]$

Hence:

$$k = \frac{J_s}{2\ \lambda_w [FeS_2]^o} \qquad (3.13)$$

The sulphate concentration in the tailings due to pyrite oxidation has been previously (Halbert et al. 1983) shown to be given by:

$$-D \frac{\partial^2 C_s}{\partial z^2} + \frac{\partial C_s}{\partial t} + u \frac{\partial C_s}{\partial z} = \frac{J_s}{\lambda_w} \exp(-kt) \qquad (3.14)$$

where

$C_s = [SO_4^{2-}]_p = C_s(t,z)$ = concentration of dissolved sulphate due to pyrite oxidation at time "t" and position "z" in the tailings ($mol.m^{-3}$)

u = porewater velocity = net precipitation (P) movement into the unsaturated zone ($m.yr^{-1}$).

D = diffusivity of sulphate in tht tailings porewater ($m^2.yr^{-1}$)

Since "z" is small (less than 1 m), dispersion can be neglected and the solution to equation (3.14) is:

$$C_s(t,z) = \frac{J_s}{\lambda_w k} \exp(-kt)\ [1-\exp(-kz/P)]$$

as: $\frac{kz}{P} << 1$

$$\frac{[1-\exp(-kz/P)]}{k} \simeq \frac{z}{P}$$

Thus:

$$C_s(t,z) = \frac{J_s}{P} \frac{z}{\lambda_w} \exp(-kt) \qquad (3.15)$$

At time t=0, and depth λ_w:

$$C_s(0,\lambda_w) = J_s/P = C_s^o$$

thus the solution to equation (3.15) can be written as:

$$C_s(t,z) = C_s^o \exp(-kt)\ (z/\lambda_w) \qquad (3.16)$$

The solution for $[Ca^{2+}]$ (equation (3.9)), becomes dependent on "t" and "z":

$$[Ca^{2+}] = C_T = \left\{ \frac{K^{o}_{s,CaSO_4}}{f^2} + 0.25\ [C^{o}_{s} \exp(-kt)\ (z/\lambda_w)]^2 \right\}^{1/2} - 0.5\ [C^{o}_{s} \exp(-kt)\ (z/\lambda_w)] \qquad (3.17)$$

3.4 Determination of Mole Fraction of Radium in Reprecipitated Gypsum

Determination of the mole fraction of radium in the solid phase requires consideration of several physical/chemical processes occurring in the unsaturated zone. Only after the behaviour of gypsum is modelled can the radium content of the gypsum be determined.

In pyritic tailings, it is not uncommon to observe an increase in sulphate concentration with depth. Due to this phenomenon some of the gypsum which dissolves in the upper portions of the unsaturated zone will reprecipitate (as the $[SO_4^{2-}]$ increases) in the lower portions of the zone. Thus the unsaturated zone can be thought to consist of an upper layer, in which all the gypsum has dissolved and migrated downward with the porewater, and a lower layer in which some of the gypsum has reprecipitated. These will be referred to as the gypsum-depleted layer and the gypsum-enriched layer in the following discussion. The relative heights of these two layers will change with time; the depeleted zone increasing until the entire unsaturated zone is depleted of gypsum. After this time, a number of physical and chemical processes can be ignored and the source term model becomes simpler.

3.4.1 Depth of Depleted Zone and Time to Total Depletion

As shown in Figure A-1, let "h" be the depth of the gypsum-depleted zone at time t_h. Thus, the amount of gypsum dissolved in t_h per unit surface area (i.e. the mass/area of dissolved gypsum out of the depleted zone) is:

$$\text{mass/area} = g_o h \tag{3.18}$$

where

g_o = initial concentration of gypsum (solid) ($\mathrm{mol.m^{-3}}$)
h = depth of gypsum depleted zone (m)

By a simple material balance, this must be equal to the time integral of the flux of dissolved gypsum into the enriched zone. That is:

$$g_o h = P \int_o^{t_h} C_T(t,h)\,dt \tag{3.19}$$

where

P = net precipitation ($\mathrm{m.yr^{-1}}$)
$C_T(t,h)$ = dissolved gypsum concentration at time t, depth h ($\mathrm{mol.m^{-3}}$)

Equation (3.18) is the basis for many of the subsequent calculations. By incorporating equation (3.17), the integral on the right hand side of equation (3.19) can be written as:

$$\mathrm{RHS} = P \;{}^{t_h} \left[\left\{ \frac{K^o_{s,CaSO_4}}{f^2} + 0.25 \left(C^o_s \exp(-kt_h)\,(h/\lambda_w) \right)^2 \right\}^{1/2} - 0.5\,[C^o_s \exp(-kt_h)\,(h/\lambda_w)] \right] dt$$

This integral becomes very difficult to solve analytically and the details will not be provided here. In short, equation (3.19) can be written as:

$$g_o h = Pat + \frac{P}{k}\left\{\left[a^2+\frac{(C_s^o)^2}{4}(h/\lambda_w)^2\right]^{1/2} - \left[a^2+\frac{(C_s^o)^2}{4}\exp(-2kt)\ (h/\lambda_w)^2\right]^{1/2}\right\}$$

$$-\frac{P}{2k}C_s^o\ [1-\exp(-kt)]\ (h/\lambda_w) - \frac{Pa}{k}\ln\left\{\frac{a+\left[a^2+\frac{(C_s^o)^2}{4}(h/\lambda_w)^2\right]^{1/2}}{a+\left[a^2+\frac{(C_s^o)^2}{4}\exp(-2kt)\ (h/\lambda_w)^2\right]^{1/2}}\right\} \quad (3.20)$$

Note: $a = \frac{\sqrt{K^o_{s,CaSO_4}}}{f}$

From this expression, the depth of the depleted zone (h) can be determined for any time t_h by an iterative (i.e. trial and error) solution.

By letting $h = \lambda_w$ in equation (3.20) the time for the entire unsaturated zone to be depleted of gypsum (t_f) can be determined. This also requires an iterative solution of:

$$g_o h = Pat_f + \frac{P}{k}\left\{\left[a^2+\frac{(C_s^o)^2}{4}\right]^{1/2} - \left[a^2+\frac{(C_s^o)^2}{4}\exp(-2kt_f)\right]^{1/2}\right\}$$

$$-\frac{P}{2k}C_s^o\ [1-\exp(-kt_f)] - \frac{Pa}{k}\ln\left\{\frac{(a+\left[a^2+\frac{(C_s^o)^2}{4}\right]^{1/2}}{a+\left[a^2+\frac{(C_s^o)^2}{4}\exp(-2kt_f)\right]^{1/2}}\right\} \quad (3.21)$$

3.4.2 Reprecipitated Gypsum Content (Enriched Layer)

Having constructed the mass balance in equation (3.19) and developed the solution in equation (3.20), determination of the reprecipitated gypsum content in the enriched layer is relatively straight forward. A gypsum mass balance on the layer λ_w-h (see Figure A-1) would take the form:

$$\begin{bmatrix}\text{reprecipitated}\\ \text{gypsum content}\\ \text{in } (\lambda_w-h)\end{bmatrix} = \begin{bmatrix}\text{time integral}\\ \text{of gypsum flux}\\ \text{into } (\lambda_w-h)\end{bmatrix} - \begin{bmatrix}\text{time integral}\\ \text{of gypsum flux}\\ \text{out (i.e. at } \lambda_w)\end{bmatrix}$$

That is, the difference between the flux of dissolved gypsum into and out of the layer must be retained within the layer as reprecipitated gypsum. This can be expressed as:

$$m(t, \lambda_w-h) = g_o h - P \int_o^{t_h} C_T(t, \lambda_w)\, dt$$

where

$m(t,\lambda_w-h)$ = reprecipitated gypsum content per unit area ($mol.m^{-2}$)

Substituting the expression previously determined for the integral on the right hand side:

$$m(t,\lambda_w-h)=g_o h - Pat - \frac{P}{k}\left\{\left[a^2 + \frac{(C_s^o)^2}{4}\right]^{1/2} - \left[a^2 + \frac{(C_s^o)^2 \exp(-2kt)}{4}\right]^{1/2}\right\}$$

$$+ \frac{P}{2k} C_s^o \left[1-\exp(-kt)\right] + \frac{Pa}{k} \ln\left\{\frac{a + \left[a^2 + \frac{(C_s^o)^2}{4}\right]^{1/2}}{a + \left[a^2 + \frac{(C_s^o)^2 \exp(-2kt)}{4}\right]^{1/2}}\right\} \quad (3.22)$$

Note that the concentration of reprecipitated gypsum is easily determined by dividing $m(t,\lambda_w-h)$ by the depth of the enriched layer (λ_w-h).

That is:

$$[gyp]_R = \frac{m(t,\lambda_w-h)}{\lambda_w-h} \quad \text{for } h< \lambda_w$$

$$= 0 \qquad \text{for } h\geq \lambda_w$$

where:

$[gyp]_R$ = concentration of reprecipitated gypsum $(mol.m^{-3})$

3.4.3 Radium Content of Reprecipitated Gypsum

Again, the calculations are based on a relatively simple material balance; on radium in this instance. Consider the change in depth of the depleted zone (Δh) with a change in time (Δt). Then,

$$\begin{bmatrix}\text{dissolved radium}\\ \text{input in } \Delta t\end{bmatrix} = \begin{bmatrix}\text{radium reprecipitated}\\ \text{in } \Delta t\end{bmatrix} - \begin{bmatrix}\text{radium leaving}\\ \text{with porewater in } \Delta t\end{bmatrix}$$

or

$$g_o\ \overline{X}^o_{Ra/gyp}\ \Delta h + PC_{Ra}\ \Delta t = (m_{t+\Delta t}x_{t+\Delta t}-m_t x_t) + P\int_t^{t+\Delta t} C_T y\,dt \qquad (3.23)$$

where

$\overline{X}^o_{Ra/gyp}$ = original mol fraction of radium in gypsum solids

C_{Ra} = $C_{Ra}(t)$ = concentration of radium due to 'other sources' (eg. ingrowth from thorium which is explained and calculated later) $(mol.m^{-3})$

y = mol fraction of radium in the liquid phase

The integral on the right hand side of equation (3.23) can be written as:

$$\int_{t}^{t+\Delta t} PC_{T}y\ dt = PK\left\{\frac{C_{T,t+\Delta t}\ x_{t+\Delta t} - C_{T,t}x_{t}}{\ln\left(\dfrac{C_{T,t+\Delta t}x_{t+\Delta t}}{C_{T,t}x_{t}}\right)}\right\}\Delta t$$

And again, one is left with an equation which requires a trial and error solution for $x_{t+\Delta t}$, namely:

$$g_{o}\bar{x}^{o}_{Ra/gyp}\ \Delta h+PC_{Ra}\Delta t = m_{t+\Delta t}x_{t+\Delta t}-m_{t}x_{t}+(PK)\left\{\frac{C_{T,t+\Delta t}x_{t+\Delta t} - C_{T,t}x_{t}}{\ln\left(\dfrac{C_{T,t+\Delta t}x_{t+\Delta t}}{C_{T,t}x_{t}}\right)}\right\}\Delta t \qquad (3.24)$$

The calculated 'x' represents the mol fraction of radium associated with reprecipitated gypsum. Although this is not strictly equal to the mol fraction of radium associated with all of the 'gypsum-like' solids, it is sufficient for our purposes. This arises since the liquid phase (i.e. $[Ra^{2+}]$ in porewater) is in equilibrium with the reprecipitated gypsum.

Thus, recalling equation (3.2):

$$[Ra^{2+}(aq)] = Kx[gyp(aq)]$$

all the unknown quantities have been explicitly defined and hence the $[Ra^{2+}(aq)]$ can be calculated. As mentioned previously, equation (3.2) is valid only for times less than t_f (the time to totally deplete the unsaturated zone of gypsum). Once the zone has been depleted of gypsum, the dissolved radium from 'other sources' provides the radium content in the porewater. That is:

$$[Ra(aq)] = C_{Ra} \quad \text{for } t \geq t_f \qquad (3.25)$$

4.0 POREWATER pH AS A FUNCTION OF TIME

The pH equilibrium model is based on a quartz/kaolinite/ferric hydroxide/gypsum/sulfuric acid system. The following equilibrium reactions (Stumm and Morgan 1981), in the unsaturated zone were considered:

Equilibrium Reactions in the Unsaturated Zone	Log K
$SiO_2 + 2H_2O \rightleftarrows H_4SiO_{4(s)}$	-2.7
$3H^+ + 0.5\ Al_2Si_2O_5(OH)_{4(s)} + 0.5H_2O \rightleftarrows Al^{3+} + H_4SiO_4$	3.8
$Al^{3+} + H_2O \rightleftarrows Al(OH)^{2+} + H^+$	-4.97
$Al^{3+} + 2H_2O \rightleftarrows Al(OH)_2^{\ +} + 2H^+$	-9.30
$Al^{3+} + 3H_2O \rightleftarrows Al(OH)_3^{\ \circ}(aq) + 3H^+$	-15.00
$Al^{3+} + SO_4^{\ 2-} \rightleftarrows AlSO_4^{\ +}$	3.20
$Fe(OH)_3(s) + 3H^+ \rightleftarrows Fe^{3+} + 3H_2O$	5.97 (K^c)
$Fe(OH)_{3(s)} + 2H^+ \rightleftarrows Fe(OH)^{2+} + 2H_2O$	3.78 (K^c)
$Fe(OH)_{3(s)} + H^+ \rightleftarrows Fe(OH)_2^{\ +} + H_2O$	0.30 (K^c)
$Fe^{3+} + SO_4^{\ 2-} \rightleftarrows FeSO_4^{\ +}(aq)$	3.02
$CaSO_4(s) \rightleftarrows Ca^{2+} + SO_4^{\ 2-}$	-4.6
$H_2CO_3 \rightleftarrows H^+ + HCO_3^{\ -}$	-6.5
	$pCO_2 = -3.5$
	$K_H = -1.27$

(K^c values are experimental from tailings oxidation studies.)

An expression implicitly relating pH to total dissolved calcium (C_T) and total dissolved sulphate (C_T+C_S) can be derived given the above. Again, the algebra is rather involved and tedious but yields the following expression.

$$2C_s + 3\left(\frac{10^{6.5}}{f^{\ 9/4}} + 10^{5.97}\right)10^{-3pH} + 2\left(\frac{10^{1.53}}{f} + 10^{3.78}\right)10^{-2pH} + 10^{0.3-pH}$$

$$= \left[\frac{(C_s+C_T)\ [2 - f^3(10^{9.7-3pH}) - (10^{8.99-3pH})]}{1 + f^3(10^{9.7-3pH}) + f^{3/4}(10^{8.99-3pH})}\right] + \left[\frac{10^{pH-11.3}}{f^{1/4}}\right] \quad (4.1)$$

As before, an iterative solution is required, resulting in a pH profile over time and depth in the tailings.

5.0 DETERMINATION OF THORIUM CONCENTRATION AND RADIUM INGROWTH FROM THORIUM AS FUNCTIONS OF TIME

5.1 Thorium Concentration

5.1.1 In Solid Phase

The amorphous fraction of the thorium-bearing material ThO_2 is assumed to be dissolved in an exponential fashion. Once this is all dissolved, a certain fraction of the thorium will remain in its crystalline state. Thus:

$$[Th_s] = X_{Th}[Th]^o + (1-X_{Th})\ [Th]^o \exp(-k_{Th}t) \quad (5.1)$$

where

$[Th_s]$ = $[Th(t)_s]$ = time-varying concentration of Th in the solid phase ($mol.m^{-3}$)

$[Th]^o$ = initial concentration of total thorium in the solid phase ($mol.m^{-3}$)

k_{Th} = arbitrary constant representing the rate of amorphous Th dissolution (yr^{-1}).

5.1.2 In Liquid Phase

The rate of disappearance of amorphous Th from the solid phase must be equal to the rate of its appearance in the liquid (porewater) phase. Thus, the time-varying concentration of total Th in porewater at the bottom of the saturated zone ($[Th]_{\lambda_w}$ $mol.m^{-3}$) is given by:

$$[Th_L]_{\lambda_w} = \lambda_w k_{Th} (1-X_{Th})[Th]^o \exp(-k_{Th}t)/ P \qquad (5.2)$$

5.2 Radium Ingrowth from Thorium

The ingrowth of radium from thorium can be estimated from the general laws of radioactive decay. The decay of radium can be represented by:

$$A \xrightarrow{\lambda_A} \qquad B \xrightarrow{\lambda_B}$$

where A and B represent the number of atoms of thorium and radium which are present at any time t, and λ_A and λ_B are the decay constants of thorium and radium respectively. If the only source of thorium atoms is the initial supply $A = A_o$ at $t=0$ and if $B = 0$ at $t = 0$, then the rate of change in the number of radium atoms is given by (Evans 1955):

$$\frac{dB}{dt} = A_o \lambda_A \exp(-\lambda_A t) - B\lambda_B$$

The solution of this differential equation is :

$$B(t) = A_o \left(\frac{\lambda_A}{\lambda_B - \lambda_A}\right)\left([\exp(-\lambda_A t) - \exp(-\lambda_B t)]\right) \qquad (5.3)$$

which is an expression for the number of radium atoms, B(t), at any time t.

5.2.1 Liquid Phase

The residence time of the porewater is small relative to the half-lives of thorium and radium. For such relatively short times, $\lambda_A t \rightarrow 0$ and $\lambda_B t \rightarrow 0$, and equation (5.3) simplifies to:

$$B = \lambda_A A_o t$$

or specifically, in the present case,

$$[Ra\text{-}226] = \lambda_{Th\text{-}230} [Th\text{-}230]^o t \qquad (5.4)$$

Equation (5.4) gives the total number of atoms or amount of radium ingrown from thorium in time t. To determine the amount of radium in the liquid phase only, that fraction of the ingrown radium which is assumed to be soluble must be considered. As well, the time-varying concentration of Th-230 must be expressed in terms of the (previously derived) time-varying concentration of the total Th.

Thus:

$$[Ra_L]_{\lambda w} = (1-X_{Ra/Th}) \left\{ \frac{\lambda_{Th} \ (X_{Th/Th}[Th_s])}{P/\lambda_w} \right\} \qquad (5.5)$$

where

$[Ra_L]_{\lambda w}$ = $[Ra(t)_L]_{\lambda w}$ = concentration of Ra-226 in the liquid phase derived from ingrowth to Th-230 at λ_w at any time t ($mol.m^{-3}$)

5.2.2 In Solid Phase

Calculation of the Ra concentration in the solid phase employs equation (5.3) as well as assumptions outlined earlier. A further assumption is that the amorphous fraction of the total Th does not contribute to solid phase radium. This is justified on the grounds that the time required for removal (by leaching) of the amorphous thorium (200-300 years) is much less than the rate of ingrowth of Ra. It is further assumed that some of the Ra ingrowth to the Th matrix will dissolve and thus not contribute to the solid phase Ra concentration. Thus, substituting thorium-230 and radium-226 for A and B respectively in equation (5.3) and replacing the $[\text{Th-230}]^{o}$ term in equation (5.3) is replaced by X_{Th} $[\text{Th-230}]^{o}$, we get:

$$[Ra_s]_{Th} = X_{Ra/Th} \left\{ (X_{Th}[\text{Th-230}]^{o} \left[\frac{\lambda_{Th}}{\lambda_{Ra} - \lambda_{Th}}\right] \left[\exp(-\lambda_{Th}t) - \exp(-\lambda_{Ra}t)\right] \right\} \quad (5.6)$$

6.0 DETERMINATION OF RADIUM ASSOCIATED WITH THE SOLID PHASE AS A FUNCTION OF TIME AND DEPTH

Radium associated with the solid phase will be attributable to three main sources:

$[Ra_s]_m$ = insoluble radium associated with host tailings material ($mol.m^{-3}$)

$[Ra_s]_{Th}$ = radium ingrowth to the thorium matrix from thorium-230 decay ($mol.m^{-3}$)

$[Ra_s]_{gyp}$ =radium associated with solid gypsum ($mol.m^{-3}$)

Thus, the concentration of radium in the solid phase will differ between the gypsum-depleted layer and the gypsum-enriched layer, by the amount of radium associated with the solid gypsum phase.

For the gypsum-depleted layer, the radium content of the solid phase can be determined from:

$$[Ra_s]_1 = [Ra_s]_m^{o} \ \exp(-\lambda_{Ra}t) + [Ra_s]_{Th} \qquad (6.1)$$

where:

$[Ra_s]_1$ = concentration of radium in the solid phase in the depleted zone (layer 1) ($mol.m^{-3}$)

$[Ra_s]^o_m$ = initial radium concentration in host tailings material ($mol.m^{-3}$)

λ_{Ra} = decay constant for Ra-226 (= $\ln 2/t_{1/2}$) (yr^{-1})

$[Ra_s]_{Th}$ = concentration of solid phase radium due to ingrowth from thorium from equation (5.6) ($mol.m^{-3}$)

For the gypsum-enriched layer, the analagous equation is:

$$[Ra_s]_2 = [Ra_s]_1 + \bar{X}^o_{Ra/gyp}\, g_o + \frac{xm}{(\lambda_w - h)} \qquad (6.2)$$

where:

$[Ra_s]_2$ = concentration of Ra in the solid phase in the enriched zone (layer 2) ($mol.m^{-3}$)

$\bar{X}^o_{Ra/gyp}$ = initial mol fraction of Ra in gypsum

g_o = initial gypsum concentration in tailings ($mol.m^{-3}$)

x = mol fraction of radium in reprecipitated gypsum

m = reprecipitated gypsum content ($mol.m^{-2}$)

$(\lambda_w - h)$ = depth of enriched zone (m)

It should be noted that equation (6.2), for the enriched zone, is valid only for times less than t_f (i.e. the time to entirely deplete the unsaturated zone of gypsum). After this time, the unsaturated zone can be considered homogenous with respect to $[Ra_s]$ with equation (6.1) remaining operative.

7.0 EFFECTS OF THE MIXING ZONE

As mentioned earlier, the upper 1 m of the saturated zone is modelled as a mixing zone. Different chemical species and radionuclides will be affected differently by this mixing zone.

7.1 Ionic Concentrations and pH Calculation

7.1.1 Calcium, Sulphate and Iron Concentrations

For all times less than t_f (the time to deplete gypsum in the unsaturated zone), the concentrations of $SO_4^{2-} + Ca^{2+}$ in the porewater exiting the unsaturated zone (i.e. at λ_w) will be at saturation levels. Thus, no net dissolution of gypsum in the mixing zone will occur for times less than t_f and the $Ca^{2+} + SO_4{}^{2-}$ concentrations will be equal to those at λ_w. That is:

$$[SO_4^{2-}]_{mix} = [SO_4{}^{2-}]_{\lambda_w} \qquad \text{for } t \leq t_f \qquad (7.1)$$

$$[Ca^{2+}]_{mix} = [Ca^{2+}]_{\lambda_w} \qquad \text{for } t \leq t_f$$

After t_f, $[Ca^{2+}]_{\lambda_w}$ equals zero (by definition), and the $[SO_4^{-2}]_{\lambda_w}$ is also greatly decreased. Thus dissolution of gypsum in the mixing zone begins to occur. From equation (3.6):

$$\frac{K^{o}_{s,CaSO_4}}{f^2} = [SO_4{}^{2-}]\,[Ca^{2+}]$$

In the mixing zone

$$[SO_4^{2-}]_{mix} = [SO_4{}^{2-}]_{\lambda_w} + [SO_4{}^{2-}]_{diss}$$

$$[Ca^{2+}]_{mix} = [Ca^{2+}]_{diss}$$

where

$[SO_4{}^{2-}]_{diss} = [Ca^{2+}]_{diss}$ = concentration of ions derived from gypsum dissolution in mixing zone = C

Substituting into equation (3.6) an expression can derived for C:

$$C = -1/2\ [SO_4^{2-}]_{\lambda_w} + 1/2 \sqrt{[SO_4^{2-}]^2_{\lambda_w} + 4\ K^{o}_{s,CaSO_4}/f^2} \quad (7.2)$$

The initial inventory of gypsum is calculated using quantities previously derived:

$$M^{o}_{gyp} = g_o\ d_{mix} \quad (7.3)$$

where

g_o = initial gypsum inventory ($mol.m^{-3}$)

d_{mix} = depth of the mixing zone (1 m)

M^{o}_{gyp} = mol of gypsum per unit surface area initially present the in mixing zone ($mol.m^{-2}$)

A careful mass balance must be maintained to ensure that no more gypsum than this initial amount is dissolved over time in the mixing zone.

The cumulative amount dissolved to the end of the current time period is given by:

$$M_{gyp,n} = M_{gyp,n-1} + \left[\frac{(C_n + C_{n-1})}{2}\right] P\,\Delta t \quad (7.4)$$

where:

$M_{gyp,n}$ = mol of gypsum dissolved to end of time period n per unit area ($mol.m^{-2}$)

$M_{gyp,n-1}$ = mol of gypsum dissolved to end of time period (n-1) per unit area ($mol.m^{-2}$)

C_{n-1}, C_n = $[Ca^{2+}]_{diss} = [SO_4^{2-}]_{diss}$ in time periods (n-1) and n respectively ($mol.m^{-3}$)
P = net precipitation ($m.yr^{-1}$)
Δt = length of time period (yr)

For $M_{gyp,n} \leq M^{o}_{gyp}$

$$[Ca^{2+}]_{mix} = [Ca^{2+}]_{diss} = C$$

$$[SO_4^{2-}]_{mix} = C + [SO_4^{2-}]_{\lambda_w} \quad \text{for } t > t_f$$

For $M_{gyp,n} > M^{o}_{gyp}$ (7.5)

$$[Ca^{2+}]_{mix} = 0$$

$$[SO_4^{2-}]_{mix} = [SO_4^{2-}]_{\lambda_w}$$

Iron transferred into the mixing zone ($[Fe^{3+}]_{mix}$) is calculated as a function of pH at λ_w:

$$[Fe^{3+}]_{mix} = 10^{5.97-3pH} + \left\{ \frac{f^{3/4}\,(10^{8.99-3pH})\,[SO_4^{2-}]_{mix}}{1 + f^{3/4}(10^{8.99-3pH}) + f^{3}(10^{9.7-3pH})} \right\}$$

$$+ 10^{3.78-2pH} + 10^{0.3-pH} \quad (pH = pH \text{ at } \lambda_w) \quad (7.6)$$

The derivation concerning the speciation of sulphate (SO_4^{2-}, $AlSO_4^{+}$, $FeSO_4^{+}$) is similar to that in the lake submodel discussed in Appendix C. Ferric iron is reduced to Fe^{2+} in the mixing zone:

$$2Fe^{3+} + FeS_2 \rightarrow 3Fe^{2+} + 2S^{o}$$

Therefore $[Fe^{2+}]_{mix} = 3/2\,[Fe^{3+}]_{mix}$ (7.7)

7.1.2 pH Calculation

The pH in the mixing zone is modelled as an explicit function of Fe^{3+}, Ca^{2+} and SO_4^{2-}.

The pH is calculated by a trial and error solution of the following equation:

$$2[Ca^{2+}]_{mix} + 2[Fe^{2+}]_{mix} + \left[\frac{3(10^{6.8-3pH})}{f^{9/4}}\right] + \left[\frac{2(10^{1.83-2pH})}{f}\right] + \left[\frac{10^{2.9-pH}}{f^{1/4}}\right]$$

$$= [SO_4^{2-}]_{mix}\left[\frac{2 - f^3\ 10^{10-3pH}}{1 + f^3\ 10^{10-3pH}}\right] + \left[\frac{10^{pH-10.47}}{f^{1/4}}\right] \qquad (7.8)$$

7.2 Radionuclide Concentrations in the Mixing Zone (Liquid Phase)

7.2.1 Radium Concentration

The differential equation for the behaviour of radium in the mixing zone can be written as:

$$\lambda_{mix}\ \frac{d}{dt}\left\{\epsilon[Ra]_{mix} + \rho_T s\right\} = P[\bar{Ra}]_{\lambda_w} - P[Ra]_{mix} \qquad (7.9)$$

where

ϵ = pore volume fraction
ρ_T = density of tailings $(g.m^{-3})$
s = surface-adsorbed radium $(mol.g^{-1})$
P = precipitation $(m.yr^{-1})$

$[\bar{Ra}]_{\lambda_w}$ = the average concentration of radium entering the mixing zone between t and t+ Δt $(mol.m^{-3})$
$[Ra]_{mix}$ = porewater concentration of radium in the mixing zone $(mol.m^{-3})$

The surface absorbed radium may be calculated from:

$$s = K_{D,Ra} \, [Ra]_{mix}$$

where

$K_{D,Ra}$ = transfer coefficient for Ra-226 from the liquid phase to the solid phase ($m^3.g^{-1}$)

On substitution into equation (7.9) the differential relationship becomes:

$$\lambda_{mix} \; \epsilon \left\{ 1 + \frac{\rho_T K_{D,Ra}}{\epsilon} \right\} \frac{d[Ra]_{mix}}{dt} = P[\bar{Ra}]_{\lambda_w} - P[Ra]_{mix} \qquad (7.10)$$

Integration of equation (7.10) yields:

$$[Ra]_{t+\Delta t} = [\bar{Ra}]_{\lambda_w} + [Ra]_t - [\bar{Ra}]_{\lambda_w} \left\{ \frac{-P \, \Delta t}{\epsilon \, \lambda_{mix} \, (1+K_{D,Ra} \; \rho_T/\epsilon)} \right\} \qquad (7.11)$$

7.2.2 Thorium Concentration

The thorium concentration in the mixing zone porewater ($[Th]_{mix}$) is calculated as a function of pH (Langmuir and Herman 1980). The predicted thorium concentration is very sensitive to the form (amorphous or crystalline) in which thorium is assumed to be present. This is reflected in the factor K_{Th} in the equation below:

$$[Th]_{mix} = \frac{10^{K_{Th} - 4pH}}{f^4} \Big\{ 1 + f^{7/4} \, (10^{pH - 3.2}) + f^3 \, (10^{2pH - 6.93})$$

$$+ f^{15/4} \, (10^{3pH - 11.7}) + f^4 \, (10^{4\,pH - 15.9})$$

$$+ f^4 \, (10^{5.45} \, [SO_4^{2-}]_{mix} \quad + f^6 \, 10^{9.73} \, [SO_4^{2-}]^2_{mix} \Big\} \qquad (7.12)$$

K_{Th} ranges from a high value of 6.3 for completely soluble (amorphous) thorium to a low of 1.78 if completely insoluble (crystalline) thorium is assumed. K_{Th} is calculated based on the factor X_{Th} described in section 2:

$$K_{Th} = (10^{6.3})^{(1-X_{Th})} (10^{1.78})^{X_{Th}} \qquad (7.13)$$

where

X_{Th} = fraction of total Th assumed to be in crystalline form (immobile)

8.0 MODEL REVISIONS FOR ALTERNATIVE SCENARIOS

The mathematical models just described, and the parameter distributions presented in Table A.1 are applicable to the reference tailings site described earlier. The models and parameters represent tailings which have not been significantly treated or modified. They thus represent a "base case" (Scenario 1) against which various management options may be evaluated. As mentioned in Volume 1 of this report, three alternative technologies were modelled during this study. The characteristics of the various technologies were tabulated in Table 2.1 of Volume 1, which is reproduced for convenience as Table A.2. The following sections detail model and parameter changes required to simulate the three alternative management options (Scenarios 2,3 and 4).

8.1 Scenario 2: Pyrite Reduction

Scenario 2 assumes that a 1 metre deep layer of depyritized tailings has been placed on the reference tailings area. The pyrite content of the layer has been reduced by about 90 percent; from 730 $mol.m^{-3}$ to 70 $mol.m^{-3}$. This scenario requires no model changes and only one parameter change, as indicated below:

change $[FeS_2]^o$ from: constant = 730 $mol.m^{-3}$
to: constant = 70 $mol.m^{-3}$

8.2 Scenario 3: Pyrite Reduction and Limestone Addition

Scenario 3 involves the pyrite reduction in, and limestone addition to, the upper layer of the tailings. The addition of limestone alters the chemical equilibria in the tailings and requires a number of model revisions. The changes required are detailed below.

When limestone ($CaCO_3$) is added to the top layer of the tailings, the following two equilibria must be satisfied simultaneously:

$$CaCO_3(s) + H^+ \rightleftarrows Ca^{2+} + HCO_3^- \qquad {}^*K_s = 10^{+2.07} \text{ (Temp=10°C)}$$
$$CaSO_4(s) \rightleftarrows Ca^{2+} + SO_4^{2-} \qquad K_{so} = 10^{-4.6} \tag{8.1}$$

This will be the case as long as both gypsum ($CaSO_4$) and limestone are present.

The solution requires some assumptions concerning the (total) carbonate content of the porewater. This carbonate content ($[CO_2]_T$) is defined as follows:

$$[CO_2]_T = [HCO_3^-] + [H_2CO_3]$$

where (8.2)

$$H_2CO_3 \rightleftarrows H^+ + HCO_3^- \qquad K_1 = 10^{-6.46} \text{ (Temp=10°C)}$$

In the tailings porewater, this total carbonate is defined as a random (input) variable, varying from 10^{-3} to 5×10^{-3} $mol.L^{-1}$ (bicarbonate alkalinity of 50 mg/L to 250 mg/L as $CaCO_3$ at pH 7 to 9) (see Stumm and Morgan (1981)).

Thus, the concentrations of bicarbonate, calcium and sulphate ions at λ_w become:

$$[HCO^-{}_3]_{calc} = [CO_2]_T \left(\frac{10^{pH}}{\dfrac{10^{+6.46}}{f^{1/4}} + 10^{pH}} \right) \qquad (8.3)$$

where:

pH = pH at λ_w

$$[Ca^{2+}]_{calc} = \frac{{}^*K_s}{f^{5/4}} \; \frac{[H^+]}{[HCO_3^-]} = \left\{ \frac{10^{2.07}}{f^{5/4}\,[CO_2]_T} \right\} \left\{ \frac{10^{6.46-2pH}}{f^{1/4}} + 10^{-pH} \right\} \qquad (8.4)$$

and

$$[SO_4^{2-}]_{calc} = \frac{10^{-4.6}}{f^2 [Ca^{2+}]_{calc}} = \frac{10^{-6.67}}{f^{3/4}} \; [CO_2]_T \left\{ \frac{1}{\dfrac{10^{6.46-2\,pH}}{f^{1/4}} + 10^{-pH}} \right\} \qquad (8.5)$$

The sulphate complexes of Al^{3+} and Fe^{3+} become:

$$[AlSO_4^{+-}] = \frac{10^{6.5-3pH}}{f^{9/4}} \; [SO_4^{2-}]_{calc}$$

$$(8.6)$$

$$[FeSO_4^+] = 10^{5.97-3pH} \; [SO_4^{2-}]_{calc}$$

Since the pH is expected to stay above 7.0, these terms are negligible. Thus Al^{3+} and Fe^{3+} will also become negligible. The ionic balance is thus given by:

$$2Ca^{2+} + 2\,AlOH^{2+} + 2\,FeOH^{2+} + Fe(OH)_2^+ = 2SO_4^{2-} + HCO_3^- \qquad (8.7)$$

where:

$$2[Ca^{2+}]_{calc} = 2\left(\frac{10^{2.07}}{f^{5/4}[CO_2]_T}\left\{\frac{10^{6.46-2pH}}{f^{1/4}} + 10^{-pH}\right\}\right)$$

$$2\left\{[AlOH^{2+}] + [FeOH^{2+}]\right\} = 2\left(\left\{\frac{10^{1.53}}{f} + 10^{3.78}\right\} 10^{-2pH}\right)$$

$$[Fe(OH)_2^+] = 10^{0.3-pH}$$

$$2[SO_4^{2-}]_{calc} = 2\left[\frac{10^{-6.67}}{f^{3/4}}\quad [CO_2]_T\left\{\frac{10^{2pH}}{\frac{10^{6.46}}{f^{1/4}} + 10^{pH}}\right\}\right]$$

and

$$[HCO_3^-]_{calc} = [CO_2]_T\left\{\frac{10^{pH}}{\frac{10^{6.46}}{f^{1/4}} + 10^{pH}}\right\}$$

The revised equation for pH at λ_w is then:

$$2\left[\frac{10^{2.07}}{f^{5/4}\,[CO_2]_T}\left\{\frac{10^{6.46-2pH}}{f^{1/4}} + 10^{-pH}\right\}\right] + 2\left[\left\{\frac{10^{1.53}}{f} + 10^{3.78}\right\}10^{-2pH}\right] + 10^{0.3-pH} =$$

$$2\left[\frac{10^{-6.67}}{f^{3/4}}\quad [CO_2]_T\left\{\frac{10^{2pH}}{\frac{10^{6.46}}{f^{1/4}} + 10^{pH}}\right\}\right] + [CO_2]_T\left\{\frac{10^{pH}}{\frac{10^{6.46}}{f^{1/4}} + 10^{pH}}\right\}$$

where:

pH = porewater pH at λ_w (assumed constant while $CaSO_4$ and $CaCO_3$ present)

f = activity coefficient for divalent ions

$[CO_2]_T$ = total carbonate, an input parameter ($mol.m^{-3}$)

Note that $[SO_4^{2-}]_{calc}$ is the sum of sulphate from pyrite oxidation and gypsum dissolution. Since the pH remains high, the reaction rate ($\bar{R}_s$) is a fraction (approximately 1/20) of the rate at low pH. Furthermore, oxygen will not be consumed in the top 1m (due to low oxidation rates). Thus, the sulphate flux (J_s^*) becomes:

$$J_s^* = \frac{1}{20} \bar{R}_s \rho_T \lambda_w \tag{8.8}$$

and the sulphate concentration due to pyrite oxidation (C_s) becomes:

$$C_s(t,z) = \frac{J_s^*}{P} \frac{z}{\lambda_w} \exp(-kt) \tag{8.9}$$

Hence:

$$[SO_4]_{gypsum} = [SO_4^{2-}]_{calc} - C_s$$

Equation (3.19) becomes

$$g_o h = P \int_o^{t_h} \left\{ [SO_4^{2-}]_{calc} - \frac{J_s^*}{P} \frac{h}{\lambda_w} \exp(-kt) \right\} dt$$

or, on integration:

$$g_o h = P \left\{ [SO_4^{2-}]_{calc} t - \frac{J_s^*}{P} \frac{h}{\lambda_w k} (1 - \exp(-kt)) \right\} \tag{8.10}$$

Thus the depth of the depleted layer (h) at any time (t) and the time required for gypsum depletion (t_f) can be determined by solving equation (8.10) as appropriate.

The reprecipitated gypsum content (m) can also be calculated from:

$$m(t, \lambda_w - h) = g_o h - P \left\{ [SO_4^{2-}]_{calc} t - \frac{J_s^*}{P} \frac{1}{k} (1-\exp(-kt)) \right\} \tag{8.11}$$

The average gypsum concentration in the enriched layer is thus given by: (8.12)

$$[\mathrm{gyp}]_R = \frac{m(t,\lambda_w - h)}{\lambda_w - h}$$

The remainder of the model is the same. The major change is thus a simpler integral, since the pH was assumed constant for as long as $CaSO_4$ and $CaCO_3(s)$ are present.

The previous modifications are valid as long as both gypsum and limestone are present (i.e. up to t_f). Once gypsum is depleted, calcium (and carbonate) must satisfy only the calcite equilibrium, i.e.

$$CaCO_3(s) + H^+ \rightleftarrows HCO_3^- + Ca^{2+} \qquad {}^{*}K_s = 10^{2.07} (\mathrm{Temp}=0°\mathrm{C})$$

Since the only remaining source of sulphate is pyrite oxidation, the concentration of sulphate at $t>{}^{t}f$ is given by:

$$[SO_4^{2-}]_{calc} = \frac{J_s^* \exp(-kt)}{P} \qquad (8.11)$$

The concentration of Ca^{2+} remains constant (as calculated previously). The pH equation is modified to reflect the change in sulphate concentration:

$$2\left(\frac{10^{2.07}}{f^{5/4}\,[CO_2]_T}\left\{\frac{10^{6.46-2pH}}{f^{1/4}} + 10^{-pH}\right\}\right) + 2\left(\left\{\frac{10^{1.53}}{f} + 10^{3.78}\right\} 10^{-2pH}\right)$$

$$+ 10^{0.3-pH} = \frac{J_s^* \exp(-kt)}{P} + [CO_2]_T \left\{\frac{10^{pH}}{\frac{10^{6.46}}{f^{1/4}} + 10^{pH}}\right\} \qquad (8.12)$$

A mass balance must be maintained on limestone, in order to determine the time of limestone depletion. The initial inventory of limestone is given by:

$$[CaCO_3]^o = X_{Lime} \quad \rho_T/100 \qquad (8.13)$$

where:

$[CaCO_3]^o$ = initial inventory of limestone ($mol.m^{-3}$)

X_{lime} = initial mass fraction of $CaCO_3$ in tailings

ρ_T = tailings bulk density ($g.m^{-3}$)

100 = molecular weight of $CaCO_3$ ($g.mol^{-1}$)

The flux of dissolved limestone (J_{lime}) is calculated as:

$$J_{lime} = \left\{ [Ca^{2-}]_{calc} - [SO_4^{2-}]_{calc} \right\} P + J_s^* \exp(-kt)$$

The moles of limestone dissolved at any time can be calculated as:

$$L(t) = \left\{ [Ca^{2+}]_{calc} - [SO_4^{2-}]_{calc} \right\} Pt + \frac{J_s^*}{k} (1-\exp(-kt))$$

where:

$L(t)$ = amount of limestone dissolved in time t ($mol.m^{-2}$)

Thus, the limestone is depleted when $L(t) \geq [CaCO_3]^o$. Once the limestone is depleted, all the mathematical relationships described previously are again valid.

In summary, the input parameter changes, and additions for Scenario 3 are as follows:

change:

$[FeS_2]^o$ from: constant = 730 $mol.m^{-3}$

to: constant = 70 $mol.m^{-3}$

add:

X_{lime} = initial mass fraction of limestone in tailings

= constant = 0.01

$[CO_2]_T$ = total carbonate content ($mol.L^{-1}$)
= uniform distribution; min = 1×10^{-3}
max = 5×10^{-3}

8.3 Scenario 4: Pyrite Radionuclide Reduction and Limestone Addition

This scenario is very similar to Scenario 3; in that pyrite is reduced in, and limestone is added to the upper layer of the tailings. In addition, a large fraction of the radionuclides (Ra and Th) are removed and the gypsum content is significantly reduced. All of the model and parameter revisions and additions for limestone addition described under Scenario 3 apply to Scenario 4. In addition, a number of revisions to input parameters are required:

change:

X_{gyp} from: constant = .05
to: constant = 0.005

Ra_s^o from: triangular distribution
min = 5.0; mode = 11; max = 15
to: triangular distribution
min = 1.5; mode = 3.3; max = 4.5

$X_{U/Th}$ from: constant = 3
to: constant = 10

$X_{Ra/gyp}$ from: triangular distribution
min = 0.05; mode = 0.2; max = 0.6
to: triangular distribution
min = 0.02; mode = 0.07; max = 0.2

REFERENCES

Bjerrum, J., Schwarzenbach, G., and Sillen, L.G., 1958. "Stability Constants, Part II: Inorganic Ligands." London: The Chemical Society, Special Publication No. 7, Burlington House, W.I.

Evans, R.D., 1955. "The Atomic Nucleus." McGraw Hill Book Company Inc.

Gresens, R.L., 1981. "The Aqueous Solubility Product of Solid Solutions." Chem. Geol. Vol. 32, pp. 73-86.

Halbert, B.E., Scharer, J.M., Knapp, R.A., and Gorber, D.M., 1983. "Determination of Acid Generation in Pyritic Mine Tailings." Presented at 56th Annual Conference of WPCF, Atlanta, Georgia.

Langmuir, D. and Herman, J.S., 1980. "The Mobility of Thorium in Natural Waters at Low Temperatures." Geochim. Cosmochim. Acta., Vol. 42, pp. 547-569.

Lippmann, F., 1977. "The Solubility Product of Complex Minerals, Mixed Crystals, and Three-layer Clay Minerals." N. J. Miner. Abh., Vol. 130, p. 243.

Lippmann, F., 1980. "Phase Diagrams Depicting Aqueous Solubility of Binary Mineral Systems." N. Jl. Miner Abh., Vol. 130, pp. 1-25.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada - Phase I." Research Report Prepared for the National Uranium Tailings Program, CANMET, Energy, Mines and Resources Canada.

Stumm, W. and Morgan, J.J., 1981. "Aquatic Chemistry." John Wiley and Sons, New York, NY.

KEY TO TABLE A.1

Distribution Type	Distribution Specifications
C Constant	single numerical value
LN Lognormal	geometric mean, geometric standard deviation, minimum, maximum
N Normal	arithmetic mean, arithmetic standard deviation, minimum, maximum
T Triangular	minimum, mode (peak), maximum
U Uniform	minimum, maximum
LN Log Uniform	minimum, maximum

Table A.1

Water Source Model Parameters

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
ρ_T	tailings bulk density	$g.m^{-3}$	$1.4x10^6$	$N(1.4x10^6, 0.1x10^6, 1.1x10^6, 1.7x10^6)$	INPUT	RHOTAL
$Ra^o_{(s)}$	initial solid phase activity of Ra-226 in tailings	$Bq.g^{-1}$	11.0	T(5.0,11.0,15.0)	INPUT	CONCRA
X_{gyp}	mass fraction of gypsum in tailings	-	0.05	U(0.05,0.06)	INPUT	FRACGP
$[FeS_2]^o$	initial concentration of pyrite in tailings	$mol.m^{-3}$	730.0	Constant (730)	INPUT	PYRITE
$X_{U/Th}$	mass ratio of UO_2 to THO_2 in ore	-	3.0	Constant (3.0)	INPUT	UPERTH
X_{UO_2}	initial mass fraction of UO_2 per tonne of ore	$kg.t^{-1}$	2.0	Constant (1.2)	INPUT	UO2TON
$X_{Ra/Th}$	mol fraction of Ra-226 ingrowth to solid Th-230 matrix which is insoluble	-	0.9	T(.85,.9,.95)	INPUT	RAINTH
X_{Th}	fraction of total Th which is assumed to be immobile	-	0.35	T(0.2,0.3,0.4)	INPUT	THFRAC

Table A.1 (cont'd)

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$X_{Ra/gyp}$	fraction of total Ra-226 which is associated with gypsum	-	0.2	T(0.05,0.2,0.6)	INPUT	RAINGP
g_o	initial solid phase gypsum content of bulk tailings	$mol.m^{-3}$	calculated	calculated	SOURCE	CONC(1,1,1)
$[Ra_s]^o$	total initial solid phase concentration of Ra-226 in tailings	$mol.m^{-3}$	calculated	calculated	SOURCE	RADIUM
$[Ra_s]^o_m$	initial concentration of Ra-226 associated with host tailings material (assumed to be not leachable)	$mol.m^{-3}$ (tailings)	calculated	calculated	SOURCE	CONC(5,1,1)
$\overline{X}^o_{Ra/gyp}$	initial mol fraction of Ra-226 in gypsum	-	calculated	calculated	SOURCE	RAFRNI
$[Th\text{-}230_s]^o$	initial solid phase concentration of Th-230 in tailings	$mol.m^{-3}$	calculated	calculated	SOURCE	CTH230
$[Th\text{-}232_s]^o$	initial solid phase concentration of Th-232 in tailings	$mol.m^{-3}$	calculated	calculated	SOURCE	CTH232

Table A.1 (cont'd)

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$[Th_s]^o$	initial solid phase concentration of total Th	$mol.m^{-3}$	calculated	calculated	SOURCE	CONC(8,1,1)
$X_{Th/Th}$	mole fraction of Th-230 to Th-232	-	calculated	calculated	SOURCE	THINTH
$K^o_{s,CaSO_4}$	solubility equilibrium constant for gypsum	$mol^2.m^{-6}$	25.0	Constant(25.0)	INPUT	KSUBSO
f	activity coefficient (for divalent ions)	-	0.3	T(0.2,0.3,0.5)	INPUT	FACTCO
K	radium solid/liquid distribution coefficient (unsaturated zone)	-	0.05	U(.01,.1)	INPUT	DISTCO
λ_w	depth of unsaturated zone	m	1.0	Constant (1.0)	INPUT	DUNSAT
J_s	net sulphate flux	$mol.m^{-2}.yr^{-1}$	calculated	calculated	SOURCE	SFLUX
$\bar{R}_s$	sulphate production rate	$mol.g^{-1}.yr^{-1}$	$1.4x10^{-6}$	$LN(1.4x10^{-6},1.5)$ $(0.4x10^{-6}, 4.2x10^{-6})$	INPUT	SULPRD
$[O_2]$	oxygen concentration at the surface	$mol.m^{-3}$	9.04	Constant(9.04)	INPUT	O2CONC
D_{O_2}	oxygen diffusivity through tailings	$m^2.yr^{-1}$	70.0	T (30, 70, 130)	INPUT	O2DIFF

Table A.1 (cont'd)

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
K'	ratio of O_2 half saturation constant (KO_2) to $[O_2]$	-	$5x10^{-2}$	$LN(5x10^{-2},1.7)$ (0.02,0.25)	INPUT	KPRIME
P	net precipitation	$m.yr^{-1}$	0.4	N(0.4, 0.1, 0.1, 0.7)	INPUT	NPRCP
$C_s(t,z)$ $=C_s$	concentration of dissolved sulphate due to pyrite oxidation at time t and position z in the tailings	$mol.m^{-3}$	calculated	calculated	SOURCE	WCONC(3,2,2)
$[C_a^{2+}]$ $=C_T$	concentration of dissolved calcium at time t and position z in the tailings	$mol.m^{-3}$	calculated	calculated	SOURCE	WCONC(1,2,2)
h	depth of gypsum depleted zone at time t_h	m	calculated	calculated	SOURCE	DEPTH(2)
t_f	time for the entire unsaturated zone to be depleted of gypsum	yr	calculated	calculated	SOURCE	TFINAL
$m(t,\lambda_w-h)$	reprecipitated gypsum content per unit area	$mol.m^{-2}$	calculated	calculated	SOURCE	CONC(2,2,2)

Table A.1 (cont'd)

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$[Ra^{2+}(aq)]$	concentration of radium in the porewater	$mol.m^{-3}$	calculated	calculated	SOURCE	WCONC(7,2,2)
x	mole fraction of radium associated with reprecipitated gypsum at time t	-	calculated	calculated	SOURCE	FRACRA
$[Th_s]$	time varying concentration of total Th in the solid phase	$mol.m^{-3}$	calculated	calculated	SOURCE	CONC(8,2,2)
K_{Th}	rate constant for amorphous Th dissolution	-	.02	U(.01,.03)	INPUT	THDISS
$[Th(t)_L]\lambda_w$	time varying concentration of total Th in the pore water at the bottom of the saturated zone	$mol.m^3$	calculated	calculated	SOURCE	WCONC(8,2,2)
$[Ra(t)_L]\lambda_w$	concentration of Ra-226 (from Th-230 decay) in liquid phase at λ_w at any time t	$mol.m^{-3}$	calculated	calculated	SOURCE	WCONC(6,2,2)
$[Ra(t)_s]_{Th}$	solid phase Ra-226 concentration (due to Th-230 decay) at time t	$mol.m^{-3}$	calculated	calculated	SOURCE	CONC(6,1,2)

Table A.1 (cont'd)

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$[Ra_s]_m$	insoluble radium associated with host tailings material	$mol.m^{-3}$	calculated	calculated	SOURCE	CONC(5,1,2)
$[Ra_s]_{gyp}$	radium associated with solid gypsum	$mol.m^{-3}$	calculated	calculated	SOURCE	CRAG
$[Ra_s]_1$	concentration of radium in the solid phase in the depleted zone (layer 1)	$mol.m^{-3}$	calculated	calculated	SOURCE	CONC(7,1,2)
$[Ra_s]_2$	concentration of radium in the solid phase in the enriched zone (layer 2)	$mol.m^{-3}$	calculated	calculated	SOURCE	CONC(7,2,2)
C	concentration of Ca^{2+}, SO_4^{2-} derived from gypsum dissolution in mixing zone	$mol.m^{-3}$	calculated	calculated	SOURCE	ION
d_{mix}	depth of mixing zone	m	1.0	Constant (1.0)	INPUT	DMIX
M^o_{gyp}	Initial aqueous phase concentration of gypsum in mixing zone	$mol.m^{-2}$	calculated	calculated	SOURCE	MGYPI
$M_{gyp,n}$	mol gypsum dissolved to end of time period n	$mol.m^{-2}$	calculated	calculated	SOURCE	MGYP

Table A.1 (cont'd)

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$[Fe^3]_{mix}$	liquid phase concentration of Fe^{3+} in mixing zone	$mol.m^{-3}$	calculated	calculated	SOURCE	FE3
$[Fe^2]_{mix}$	liquid phase concentration of Fe^{2+} in mixing zone	$mol.m^{-3}$	calculated	calculated	SOURCE	WCON(9,4,2)
$[Ca^{2+}]_{mix}$	liquid phase concentration of Ca2+ in mixing zone	$mol.m^{-3}$	calculated	calculated	SOURCE	WCON(1,4,2)
$[SO_4^{2-}]_{mix}$	liquid phase concentration of SO_4^{2-} in mixing zone	$mol.m^{-3}$	calculated	calculated	SOURCE	WCONC(4,4,2)
ε	pore volume fraction	$m^3.m^{-3}$	0.4	T(.35,.4,.45)	INPUT	TVOID
$K_{D,Ra}$	Ra liquid-solid transfer coefficient (mixing zone)	$m^3.g^{-1}$	80×10^{-6}	LN(80×10^{-6},3.5) 1.9×10^{-6}, 3.4×10^{-3})	INPUT	KSUBDR
$[Ra]_{mix}$	liquid phase Ra-226 concentration in mixing zone	$mol.m^{-3}$	calculated	calculated	SOURCE	WCONC(7,4,2)
$[\overline{Ra}]_{\lambda_w}$	average liquid phase concentration of Ra-226 entering mixing zone in Δt	$mol.m^{-3}$	calculated	calculated	SOURCE	T10
$[Th]_{mix}$	liquid phase concentration of Th in mixing zone	$mol.m^{-3}$	calculated	calculated	SOURCE	WCONC(8,4,2)

Table A.2

Chemical and Radiological Characteristics - Uranium Tailings Surface Zone (Upper 1 m)

Scenario	FeS_2 %	$CaSO_4$ %	$CaCO_3$ %	Ra-226 Bq/g	Th-230 Bq/g
1 - Base Case	6.5	5	0	11.1	11.1
2 - Pyrite Reduction	0.6	5	0	11.1	11.1
3 - Pyrite Reduction and Limestone Addition	0.6	5	1.0	11.1	11.1
4 - Pyrite and Radionuclide Reduction and Limestone Addition	0.6	0.5	1.0	3.3	3.3

APPENDIX B

THE CALCULATION OF AIR SOURCE TERMS

TABLE OF CONTENTS

Page

1.0 THE DUST SOURCE MODEL (SUBROUTINE DUST) B-1

2.0 THE RADON SOURCE MODEL (SUBROUTINE RADON) B-3

REFERENCES B-9

APPENDIX B

THE CALCULATION OF AIR SOURCE TERMS

1.0 THE DUST SOURCE MODEL (SUBROUTINE DUST)

The empirical dust source model used in this assessment was developed by the Colorado Department of Health based on experimental work carried out at mining operations in the western United States (CDOH 1984). Detailed studies of dust dispersion from tailings in climates similar to those that occur in the Canadian Shield have not been identified. This model will have to be updated in the future to reflect the results of NUTP sponsored research in this area.

The emission rate of the ith radionuclide from the tailings surface at a specific moment in time t can be expressed as:

$$Q_i = 2.2 \times 10^{-6} \frac{S}{1.5} \frac{365-P}{235} \frac{f}{15} \frac{C_i}{\rho} A \qquad (1.1)$$

where

Q_i = $Q_i(t)$ = emission rate of the ith radionuclide at time t ($Bq.s^{-1}$)

S = silt content (%)

P = annual number of days with more than 0.25 mm of precipitation

f = percentage of the time wind speed at the mean tailings pile height exceeds 5.4 $m.s^{-1}$.

$C_i = C_i(t)$ = surficial concentration of the ith radionuclide at time t ($Bq.m^{-3}$)

$A = A(t)$ = the exposed tailing surface area at time t (m^2)

ρ = tailings bulk density ($g.m^{-3}$)

The empirical constant in the equation (1.1) (i.e. $2.2x10^{-6}$ $g.m^{-2}.s^{-1}$) was developed by the Colorado Department of Health as noted above.

The originators of this equation rate it as average-to-below-average in its accuracy of prediction, however, higher ratings are only given for models of well-controlled, consistent industrial sources.

The silt content (S), is determined by the proportion of material that passes through a 200 mesh screen (using the ASTM-C-136 method). This is assumed to be constant and equal to 30%. Other parameters of the particle size distribution are ignored.

The percentage of the time that the wind speed exceeds 5.4 $m.s^{-1}$ at the tailings height (f) is highly dependent on the time of year. Higher wind speeds are observed on the Canadian Shield during the winter months but frozen ground conditions and snow cover are assumed to reduce dust emissions to zero. Therefore, a distribution of the percentage of the time the wind speed exceeds 5.4 m/s for the six summer months was used. The chosen distribution is normal with a mean of 35, a standard deviation of 6 and a minimum and maximum of 23 and 46, respectively.

The second meteorological parameter in subroutine DUST is the number of days on which there is measurable precipitation (P). Again the winter months are excluded from the derivation of this distribution. The chosen distribution is normal with a mean of 123, standard deviation of 18 and a minimum and maximum of 84 and 156, respectively.

The net result of these assumptions with regard to meteorological parameters is to spread summer wind and precipitation out over the whole year. Thus each model trial will pick one emission rate to represent the entire year and it will pick that rate based on summer meteorology. As noted previously, the dust emission rate in winter is likely to be minimal hence this quantity is overestimated by the model as currently formulated. Ongong NUTP research in this area may suggest improvements in model formulation or parameter selection.

Another consideration pertains to the time dependency of the emission rates of the various compounds. There are three factors that control the time-varying behaviour of these emission rates:

- The surficial concentration of the ith compound (C_i) will vary as the compound is leached from or deposited to the tailings surface.
- The exposed area of the tailings (A) will decrease with the natural or forced revegetation of the exposed surface.
- The intrinsic dust emission rate will decrease as the tailings surface weathers.

The changes in surficial radionuclide concentrations are assumed to be taken care of by the source term model. While a decrease in the intrinsic dust emission rate with time is expected (as might occur for example with the gradual loss of erodible particles), the process is not well understood at present, and is ignored in this study.

With time the tailings surface area will be reduced by the encroachment of vegetation. For this study, the tailings surface area is set equal to the initial exposed area ($A(t) = A_o$). This is a conservative assumption since revegetation, which will occur naturally, is ignored.

2.0 THE RADON SOURCE MODEL (SUBROUTINE RADON)

The Phase I study (SENES 1984) describes a simple model for the simulation of the rate of release of radon from uranium mill tailings. The tailings pile is considered to be homogeneous with respect to radium concentration, moisture and particle size, and infinitely deep (i.e. greater than 3 to 4 m) relative to the diffusion length of radon. On this basis and using classical one-dimensional diffusion theory (no transport phenomena are considered), the radon flux at the surface of the tailings can be shown to be proportional to the product of radium concentration, radon emanation coefficient, tailings density, and radon diffusion length in the tailings. The diffusion length is proportional to the square root of the radon diffusion coefficient which is moisture dependent. An empirical relationship between the diffusion coefficient and moisture content as suggested by Rogers et al (1980) was used in the Phase I study. According to the model equations:

$$J = RaE\rho\lambda L \tag{2.1}$$

where

J = radon flux ($Bq.m^{-2}.s^{-1}$)
Ra = tailings radium concentration ($Bq.g^{-1}$)
E = emanation coefficient ($0 \leq E \leq 1$, dimensionless)
ρ = tailings density ($g.m^{-3}$)
λ = radon decay constant (s^{-1})
L = radon diffusion length in tailings (m)
$= \sqrt{\dfrac{D}{\lambda P}}$

and

P = porosity
D = effective bulk radon diffusion coefficient
$= 0.106 \times 10^{-4} \times P \times \exp(-0.261 \times M)$ ($m^2 \cdot s^{-1}$)
M = percent moisture content of tailings by weight
$= 100 \times \dfrac{\text{weight of water in tailings}}{\text{weight of wet tailings}}$

In the Phase I study, adjustments also are made for freezing conditions and snow cover.

The Phase II study requires a different radon source model because the tailings are no longer considered homogeneous or infinitely deep. As described in Appendix A, the tailings are modelled as a two-layered system: an upper, unsaturated layer in which the radium concentration is a function of depth, and a lower, saturated layer. A simplified schematic of this model is shown in Figure A.1.

The saturated layer is not a source of radon that can reach the tailings surface since the water effectively prevents its diffusion - the diffusion length of radon in the water-filled pores of the tailings is orders of magnitude less than the diffusion length in the unsaturated pores. The unsaturated layer can be divided into two zones as described by the water source model: an upper gypsum-depleted zone (with a reduced radium content) and a lower gypsum-enriched zone (with a radium-enriched content). The depth of the interface of the two zones depends on the chemistry of the tailings. The depth changes with time (over tens of years) until the entire unsaturated layer is gypsum-depleted.

In terms of radon release to the atmosphere, each of these zones can be considered to be homogeneous, with each zone characterized by an average radium concentration, moisture content, porosity, etc. This model for radon release is equivalent to a covered tailings model in which the tailings are of finite thickness and the cover also contains radium i.e. the "cover" is the upper zone and the "tailings" are the lower zone. The theoretical development of such a model has been carried out by Rogers et al (1980). The model accounts for the unique definition of tailings and cover in terms of parameters that affect radon release (such as radium concentration and radon diffusion length) and thus it can be used to model the total radon release from both zones of the unsaturated layer.

Denoting the upper zone of the unsaturated layer as zone 1 of thickness x_1, and the lower zone as zone 2 of thickness x_2, the general formula for the radon flux at the surface of the tailings is (following Rogers et al 1980, Chapter 4, Equation 14):

$$J = \frac{-D_1 a_1}{\cosh(a_1 x_1)} \left[\frac{[1+Z\coth(a_2 x_2)](C_o - s_1/a_1{}^2)\exp(a_1 x_1) + (s_1/a_1{}^2 - s_2/a_2{}^2)}{\tanh(a_1 x_1) + Z\coth(a_2 x_2)} \right] + D_1 a_1 (C_o - s_1/a_1{}^2) \qquad (2.2)$$

where

D = radon diffusion coefficient ($m^2.s^{-1}$)

C_o = radon concentration above the tailings surface (considered constant) ($Bq.m^{-3}$)

$a_i = \sqrt{\frac{\lambda P_i}{D_i}}$ = (radon diffusion length)$^{-1}$(m^{-1})

$$Z = \frac{D_1 a_1}{D_2 a_2}$$

$$s_i = \frac{(Ra)_i \, \rho_i \, \lambda \, E_i}{D_i}$$

and where subscript i = 1 or 2 refers to layer 1 or 2, respectively, and the other terms are as previously defined. (Note that the product of Ra and ρ is sometimes given only by one parameter when Ra is measured in units of $Bq.m^{-3}$.)

In the limit of very deep tailings ($x_2 \rightarrow \infty$), no "cover" ($x_1 = 0$)

and negligible atmospheric radon ($C_o \rightarrow 0$, which is a good approximation since it is orders of magnitude less than the radon concentration within the tailings), equation (2.2) becomes:

$$J = (Ra)_2 \, E_2 \, \rho_2 \sqrt{\frac{\lambda D_2}{P_2}} \tanh\left[\sqrt{\frac{P_2}{D_2}} \, x_2\right]$$

$$= (Ra)_2 \, E_2 \, \rho_2 \sqrt{\frac{\lambda D_2}{P_2}} \qquad (2.3)$$

Equation (2.3) is identical to equation (2.1) which was derived for the simpler, homogeneous, thick, one-layer model of Phase I.

The generality of equation (2.2) can accommodate completely different characteristics for each zone. However for the Phase II modelling exercise, only the thickness and radium concentration of each zone differ. The moisture and porosity, and hence the radon diffusion coefficient, of each layer are assumed to be identical. Under these conditions ($Z=1$ and setting $C_o=0$), equation (2.2) reduces to

$$J = -\frac{D}{a \cosh(ax_1)}\left[\frac{-[1 + \coth(ax_2)]s_1\exp(ax_1) + s_1 - s_2}{\tanh(ax_1) + \coth(ax_2)}\right] - \frac{Ds_1}{a} \qquad (2.4)$$

where $x_1 + x_2 = 1$ m, the depth of the unsaturated layer.

J is the unit flux in units of $Bq.m^{-2}.s^{-1}$. To obtain the total flux (J_T) for the tailings area, this must be multiplied by the area A(t) (For this study, the surface area is considered constant (A_o)). Thus:

$$J_T = J \, A(t) = JA_o$$

The variables for this model are given in Table B.2. As discussed, the zone thicknesses and radium concentrations are time-dependent and hence the radon flux is also time-dependent. The moisture dependence of D is calculated as previously stated. It is realized that the emanation coefficient E can also depend on the moisture content, especially at low moisture levels ($\leq$5%). However, for the present study, the moisture content is generally >5% and parameterization of the effects of moisture beyond that previously stated is not considered necessary.

The reduction in radon release rate due to freezing conditions and snow cover applied in the Phase I study (winter/summer flux ratio) is not presently included in the Phase II radon model, the focus of which is the simulation of layers of tailings materials of different characteristics.

REFERENCES

Colorado Department of Health (CDOH), 1984. "Fugitive Particulate Emissions." Internal Memo from T. Tistinic, Colorado Department of Health, July 2, 1984.

Rogers, V.C., Overmyer, R.F., Putzig, K.M., Jensen, C.M., Nielson, K.K., and Sermon, B.W., 1980. "Characterization of Uranium Tailings Cover Materials for Radon Flux Reduction." Prepared for the U.S. Nuclear Regulatory Commission by Argonne National Laboratory and Ford, Bacon and Davis Utah, Inc., NUREG/CR-1081, FBDU-218-2.

KEY TO TABLE B.1

Distribution Type		Distribution Specifications
C	Constant	single numerical value
N	Normal	arithmetic mean, arithmetic standard deviation, minimum, maximum
T	Triangular	minimum, mode (peak), maximum

Table B.1

Dust Source Model Variables

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$Q_i(t)$	emission rate of the ith radionuclide at time t	$Bq.s^{-1}$	calculated	calculated	DUST	EMRATE (I)
S	silt content of the tailings (% of particles <75 um)	%	30	C (30)	INPUT	SILT
P	annual number of days with more than 0.25 mm of precipitation	days	123	N(123, 18, 84, 156)	INPUT	FPREC
f	percentage of the time that the wind speed exceeds 5.4 m/s at the mean tailings height	-	35	N(35, 6, 23, 46)	INPUT	FWIND1
$C_i(t)$	surficial concentration of ith compound at time t	$Bq.g^{-1}$	calculated	calculated	SOURCE	CONC (7,1,2)
A(t)	exposed surface area of the tailings at time t	m^2	500,000	C (5×10^5)	INPUT	AEXPT
ρ	tailings bulk density	$g.m^{-3}$	1.4×10^6	N(1.4×10^6, 0.1×10^6 1.1×10^6, 1.7×10^6)	SOURCE	RHOTAL

Table B.2

Radon Source Term Model Variables

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
a	reciprocal of radon diffusion length of tailings	m^{-1}	calculated	calculated	RADON	DIFLEN
D	effective bulk radon diffusion coefficient of tailings	$m^2.s^{-1}$	calculated	calculated	RADON	EBKDFC
E	radon emanation coefficient of tailings	-	0.35	T(0.02, 0.2, 0.9)	INPUT	EMANAT
J(t)	radon flux	$Bq.m^{-2}.s^{-1}$	calculated	calculated	RADON	RNFLUX
M	percent moisture content by weight of tailings	-	15	T(5, 15, 25)	INPUT	MPCH20
P	porosity of tailings	-	0.4	T(0.35, 0.4, 0.45)	INPUT	TVOID
$(Ra)_i(t)$	radium concentration of layer i = 1 or 2	$Bq.m^{-3}$	calculated	initialized then calculated	SOURCE	CONC(7,X,2)
$x_1(t)$	depth of gypsum depleted zone	m	calculated	calculated	SOURCE	DETH1
$x_2(t)$	depth of gypsum enriched zone	m	calculated	calculated	SOURCE	DEPTH2
λ	radon decay constant	s^{-1}	2.1×10^{-6}	C (2.1×10^{-6})	INPUT	RNLAM
A(t)	exposed surface area of the tailings at time t	m^2	5.0×10^5	C (5.0×10^5)	INPUT	ATAIL

APPENDIX C

SURFACE WATER DISPERSION COMPONENT MODELS

TABLE OF CONTENTS

Page

1.0 INTRODUCTION C-1

2.0 LAKE RADIUM-226 LEVELS C-1

3.0 LAKE CALCIUM AND SULPHATE CONCENTRATIONS C-5

4.0 LAKE pH C-7

REFERENCES C-13

APPENDIX C - SURFACE WATER DISPERSION COMPONENT MODELS

1.0 INTRODUCTION

One of the goals established for the Phase I work on system modelling, which has been carried forward into the Phase II studies, involved estimating the dose to a reference receptor via six key pathways including consumption of contaminated drinking water and fish. The reference receptor, described in more detail in Volume 1, is situated on the edge of a reference lake approximately 2 km downstream of the reference tailings area. The hypothetical reference site was chosen to reflect the major environmental features of a Precambrian Shield setting. Contaminated surface and groundwater flows from the tailings area are intercepted in a major river which discharges into an oligotrophic lake (i.e. a low productivity lake) with a surface area of 160 ha and mean depth of 10 m. The drainage basin at the outflow of the lake equals 2000 ha as compared to the tailings area which covers 50 ha. The waters of the drainage basin feature a naturally low buffering capacity, low nutrient content and low total dissolved solids level.

The estimation of pH and constituent concentrations in the reference lake takes into account physical, chemical and biological factors. For example, radium rapidly adsorbs onto the surface of organic and inorganic matter and is removed from the aqueous phase with sedimenting solids. Calcium and sulphate on the other hand are affected primarily by dilution alone. These factors are itemized in the following sections. The variables included in the surface water component model and the probability distributions assigned to them are summarized in Table C.1

2.0 LAKE RADIUM-226 LEVELS

Radium behaviour in the lake is modelled on a mass basis to permit removal from the water column and feedback from the sediment to be followed with time. A detailed description of the concepts of the lake model is provided elsewhere (see discussion paper by Scharer and Halbert in Volume 2). The accumulation of radium in the lake water column may be represented by the following expression:

$$\frac{d\{Ra\}_w}{dt} = P\,A_T\,[Ra]_{T,t} - k_1\{Ra\}_w - \frac{1}{\tau_t}\{Ra\}_w + k_2\{Ra\}_s \qquad (2.1)$$

where

$[Ra]_{T,t}$ = concentration of radium in the contaminated surface runoff from the tailings area ($Bq.m^{-3}$)

P = net precipitation ($m.yr^{-1}$)

A_T = area of tailings (m^2)

$\{Ra\}_w$ = mass of radium in the lake water column (Bq)

$\{Ra\}_s$ = mass of radium in the exchange zone of the lake sediment (Bq)

k_1 = rate of radium removal from the lake water column to the lake sediment (yr^{-1})

k_2 = rate of radium release from the lake sediment to the lake water column (yr^{-1})

τ_t = retention time (or travel time) in the reference lake (yr).

Similarly, the changing inventory of radium in the lake sediment may be stated as follows:

$$\frac{d\{Ra\}_s}{dt} = k_1\{Ra\}_w - (k_2 + k_3)\{Ra\}_s \qquad (2.2)$$

where

k_3 = rate of radium removal (burial) to deep lake sediment (yr^{-1})

Collecting terms in the differential equations, they may be restated as follows:

$$\frac{d\{Ra\}_w}{dt} = (A_T \ P)\ [Ra]_{T,t} - (k_1 + 1/\tau_t)\ \{Ra\}_w + k_2\ \{Ra\}_s \qquad (2.3)$$

$$\frac{d\{Ra\}_s}{dt} = k_1\{Ra\}_w - (k_2 + k_3)\ \{Ra\}_s \qquad (2.4)$$

At time t=0, the initial radium content of the lake water column and sediment is:

$$\{Ra\}_{w,t=0} = 0$$
$$\{Ra\}_{s,t=0} = 0$$

The term of interest is the concentration of radium in the lake. Once the mass $\{Ra\}_w$ is obtained, the concentration is given by

$$[Ra]_{L,t} = \frac{\{Ra\}_w}{V}$$

where

V = the lake volume (m^3)

The two differential equations given above may be restated in matrix notation:

$$\begin{matrix} \dfrac{d\{Ra\}_w}{dt} = \\ \\ \dfrac{d\{Ra\}_s}{dt} = \end{matrix} \begin{bmatrix} a_{11} & -\ a_{12} & +\ a_{13} \\ & & \\ a_{21} & a_{22} & -\ a_{23} \end{bmatrix} \begin{bmatrix} 1 \\ \{Ra\}_w \\ \{Ra\}_s \end{bmatrix}$$

where

$a_{11} = A_T\ P\ [Ra]_{T,t} \qquad a_{21} = 0$

$a_{12} = k_1 + \frac{1}{\Upsilon_t} \qquad a_{22} = k_1$

$a_{13} = k_2 \qquad a_{23} = k_2 + k_3$

The general solution is:

$$\}Ra\{_{w,t+\Delta t} = K_w + A\ \exp(\lambda_1 \Delta t) + (X_t - A)\ \exp(\lambda_2 \Delta t) \qquad (2.5)$$

$$\}Ra\{_{s,t+\Delta t} = K_s + Be\ \exp(\lambda_1 \Delta t) + (Y_t - B)\ \exp(\lambda_2 \Delta t) \qquad (2.6)$$

where

$$K_w = \frac{a_{13}a_{21} + a_{11}a_{23}}{a_{12}a_{23} - a_{22}a_{13}}$$

$$K_s = \frac{a_{12}a_{21} + a_{11}a_{22}}{a_{12}a_{23} - a_{13}a_{22}}$$

$$\lambda_1 = -\ \frac{(a_{12} + a_{23}) + \sqrt{(a_{12} + a_{23})^2 + 4(a_{13}a_{22} - a_{12}a_{23})}}{2}$$

$\lambda_2 = -(a_{12} + a_{23} + \lambda_1)$

$X_t = \}Ra\{_{w,t} - K_w$

$Y_t = \}Ra\{_{s,t} - K_s$

$$A = \left[\frac{a_{13} + a_{23} + \lambda_1}{(a_{12}+a_{22}) - (a_{13}+a_{23})}\right]\left[\frac{(a_{12}+a_{22}+\lambda_2)\ X_t - (a_{13}+a_{23}+\lambda_2)Y_t}{\sqrt{(a_{12}+a_{23})^2 + 4(a_{13}a_{22}-a_{12}a_{23})}}\right]$$

and

$$B = \left[\frac{a_{12} + a_{22} + \lambda_1}{a_{13} + a_{23} + \lambda_1}\right] * A$$

3.0 LAKE CALCIUM AND SULPHATE CONCENTRATIONS

The concentration of the major water quality constituents, including calcium and sulphate, are modelled as conservative substances. The natural background levels assumed for several of the major constituents were:

calcium - 10^{-4} $\mathrm{mol.L^{-1}}$
sulphate - $10^{-4.15}$ $\mathrm{mol.L^{-1}}$
sodium - $10^{-4.1}$ $\mathrm{mol.L^{-1}}$
potassium - $10^{-4.8}$ $\mathrm{mol.L^{-1}}$
magnesium - $10^{-4.5}$ $\mathrm{mol.L^{-1}}$
chloride - $10^{-4.7}$ $\mathrm{mol.L^{-1}}$

The concentrations for calcium, sulphate, magnesium and chloride in Precambrian Shield water were obtained from water quality data summary reports prepared by the Ontario Ministry of the Environment (1972 to 1977). Since comparable data were not available for sodium and potassium, typical concentrations for these constituents were taken from Stumm and Morgan (1981).

The concentration of calcium and sulphate in the inflow water to the lake, taking into account dilution of the tailings area contaminated runoff with watershed drainage, is given by:

$$[\mathrm{Ca}]_{in,t} = \frac{50}{2{,}000} [\mathrm{Ca}]_{T,t} + 10^{-4}$$

or

$$[\mathrm{Ca}]_{in,t} = 2.5\mathrm{x}10^{-2} [\mathrm{CA}]_{T,t} + 10^{-4}$$

and

$$[\mathrm{SO_4}]_{in,t} = 2.5 \times 10^{-2} [\mathrm{SO_4}]_{T,t} + 10^{-4.15}$$

where

$[Ca]_{T,t}$ = concentration of calcium in the contaminated surface flow from the tailings area ($mol.L^{-1}$)

$[SO_4]_{T,t}$ = concentration of sulphate in the contaminated surface flow from the tailings area ($mol.L^{-1}$)

$[Ca]_{in,t}$ = concentration of calcium corrected for dilution in the inflow to the lake ($mol.L^{-1}$)

$[SO_4]_{in,t}$ = concentration of sulphate corrected for dilution in the inflow to the lake ($mol.L^{-1}$)

The concentration of either calcium or sulphate in the lake may be represented by a simple differential equation of the form:

$$V \frac{dC}{dt} = Q_{in}C_{in} - Q_{out}C \qquad (3.1)$$

If it is assumed that the flow into the lake equates approximately to the flow out of the basin and letting

$$\frac{Q}{V} = \frac{1}{\tau_t} \qquad (3.2)$$

then, integration for sulphate or calcium respectively yields:

$$[SO_4]_{L,t+\Delta t} = [SO_4]_{in,t+\Delta t} + [SO_4]_{L,t} - \left\{[SO_4]_{in,t+\Delta t}\right\} \exp \left\{\frac{-\Delta t}{\tau_t}\right\}$$

and

$$[Ca]_{L,t+\Delta t} = [Ca]_{in,t+\Delta t} + [Ca]_{L,t} - \left\{[Ca]_{in,t+\Delta t}\right\} \exp \left\{\frac{-\Delta t}{\tau_t}\right\}$$

where

V = the volume of the lake (m^3)

Q = A_wP = the flow out of the lake ($m^3.yr^{-1}$)

A_w = the area of the watershed (m^2)

P = the net precipitation on the watershed ($m.yr^{-1}$)
$[SO_4]_L$ = concentration of sulphate in lake water ($mol.L^{-1}$)
$[Ca]_L$ = concentration of calcium in lake water ($mol.L^{-1}$)

4.0 LAKE pH

To calculate the pH in the lake, it is assumed that the concentrations of iron and aluminum are in equilibrium with solid phases of ferric hydroxide ($Fe(OH)_3(s)$) and gibbsite ($Al(OH)_3(s)$). The concentrations of calcium and sulphate are modelled as described in the previous section. The concentrations of Na, K, Mg, and Cl are essentially unaffected by the tailings area drainage and are characterized by the assumed natural background levels. The waters are considered to be slightly super-saturated (Saturation Index = 0.3) with respect to dissolved carbonate in equilibrium with CO_2 in the air. The latter is very common in oligotrophic lakes on the Canadian Shield. The equilibrium reactions are as follows:

Equilibrium Reactions in Lake Water

	Log K
$Al(OH)_3(s) + 3H^+ \rightleftarrows Al^{3+} + 3\ H_2O$	8.5
$Al^{3+} + H_2O \rightleftarrows Al(OH)^{2+} + H^+$	-4.97
$Al^{3+} + 2H_2O \rightleftarrows Al(OH)_2^+ + 2H^+$	-9.30
$Al^{3+} + 3H_2O \rightleftarrows Al(OH)_3(aq) + 3H^+$	-15.00
$Al^{3+} + SO_4^{2-} \rightleftarrows AlSO_4^+$	3.2
$Fe(OH)_3(s) + 3H^+ \rightleftarrows Fe^{3+} + 3H_2O$	5.97 (K^c)
$Fe(OH)_3(s) + 2H^+ \rightleftarrows FeOH^{2+} + 2H_2O$	3.78 (K^c)
$Fe(OH)_3(s) + H^+ \rightleftarrows Fe(OH)_2^+ + H_2O)$	0.30 (K^c)
$H_2CO_3 \rightleftarrows H^+ + HCO_3^-$	-6.5
	$pCO_2 = -3.2$
	$K_H = -1.27$

The ionic charge balance for lake water is given by:

$$\begin{aligned}&2[Ca^{2+}]_{L,t} + 2[Mg^{2+}] + [Na^{+}] + [K^{+}] + 3[Fe^{3+}] + 2[FeOH^{2+}]\\ &+ [Fe(OH)_2^{+}] + [FeSO_4^{+}] + 3[Al^{3+}] + 2[AlOH^{2+}] + [Al(OH)_2^{+}]\\ &+ [AlSO_4^{+}] = 2[SO_4^{2-}]_{L,t} + [Cl^{-}] + [HCO_3^{-}] \qquad (4.1)\end{aligned}$$

Note that:

$$\begin{aligned}2[Mg^{2+}] + [Na^{+}] + [K^{+}] - [Cl^{-}] &= 2 \times 10^{-4.5} + 10^{-4.1}\\ &+ 10^{-4.8} - 10^{-4.7} = 10^{-3.86}\end{aligned}$$

and $$[HCO_3^{-}] = \frac{10^{pH-11.0}}{f^{1/4}}$$

The contribution of the remaining components of equation (4.1) to the pH equation are discussed below.

Trivalent Cations Al^{3+} and Fe^{3+}

The predicted concentration of Fe^{3+} is several orders of magnitude less than the concentration of Al^{3+} for any pH value:

$$\left[Fe^{3+}\right] = 10^{5.97-3pH}$$

$$\left[Al^{3+}\right] = \frac{10^{8.5-3pH}}{f^{9/4}}$$

In fact, the sum of the trivalent cations ($[Al^{3+}]+[Fe^{3+}]$) is equal to:

$$\frac{10^{5.97}}{10^{3pH}} + \frac{10^{8.5}}{f^{9/4}10^{3pH}} \approx \frac{10^{8.501}}{f^{9/4}10^{3pH}} \qquad 0.6 \le f \le 0.9$$

Divalent Hydroxo Complexes $Al(OH)^{2+}$ and $Fe(OH)^{2+}$

The concentration of $Fe(OH)^{2+}$ is given on the basis of solubility data from tailings oxidation tests:

$$[Fe(OH)^{2+}] = 10^{3.78-2pH}$$

Since the average activity coefficient for divalent cations is estimated to be 0.8; the activity of $Fe(OH)^{2+}$ is equal to:

$$\{Fe(OH)^{2+}\} = 0.8*10^{3.78-2pH} = 10^{3.683-2pH}$$

In the pH equation, the activities of $Al(OH)^{2+}$ and $Fe(OH)^{2+}$ can be readily combined (since both involve the same power with respect to pH) to give:

$$\{Al(OH)^{2+}\} + \{Fe(OH)^{2+}\} = \frac{10^{3.683}}{10^{2pH}} + \frac{10^{3.53}}{10^{2pH}} = \frac{10^{3.91}}{10^{2pH}}$$

It follows that the sum of concentrations of $Al(OH)^{2+}$ and $Fe(OH)^{2+}$ is given by:

$$[Al(OH)^{2+}] + [Fe(OH)^{2+}] = \frac{10^{3.91}}{f\ 10^{2pH}} = \frac{10^{3.91-2pH}}{f}$$

Monovalent Hydroxo Complexes $Fe(OH)_2{}^+$ and $Al(OH)_2{}^+$

The activity coefficients of monovalent ions ($f^{1/4}$) in the lake is always close to unity:

$f^{1/4} = 0.88$ if $f = 0.6$
$f^{1/4} = 0.97$ if $f = 0.9$

Thus the sum of $Fe(OH)_2{}^+$ and $Al(OH)_2{}^+$ equals:

$$[Fe(OH)_2^+] + [Al(OH)_2^+] = \frac{10^{0.3}}{10^{pH}} + \frac{10^{-0.8}}{10^{pH}} = \frac{10^{0.33}}{10^{pH}}$$

which may be restated as:

$$[Fe(OH)_2^+] + [Al(OH)_2^+] = 10^{0.33-pH}$$

Term for the Sulphate Concentration

The term given in equation (4.1) for sulphate is given by the sum of free sulphate and sulphato complexes:

$$[SO_4]_{L,t} = [SO_4^{2-}] + [FeSO_4^+] + [AlSO_4^+]$$

A simple comparison of the complexation reactions reveals that the concentration of $FeSO_4^+$ is negligible in comparison to either $AlSO_4^+$ or SO_4^{2-}. The formation of the $AlSO_4^+$ complex can be expressed from the following two equilibrium expressions:

$$Al(OH)_3(s) + 3H^+ \rightleftharpoons Al^{3+} + 3H_2O \qquad \log K = 8.5$$
$$Al^{3+} + SO_4^{2-} \rightleftharpoons AlSO_4^+ \qquad \log K = 3.2$$

Therefore:

$$Al(OH)_3(s) + 3H^+ + SO_4^{2-} \rightleftharpoons AlSO_4^+ + 3H_2O$$

The concentration of $AlSO_4^+$ is given (including the activity coefficients) as:

$$[AlSO_4^+] = \frac{f}{f^{1/4}} \; 10^{(8.5+3.2)} \; 10^{-3pH} \; [SO_4^{2-}]$$

$$= f^{3/4} \; 10^{11.7-3pH} \; [SO_4^{2-}]$$

The total concentration of sulphate includes the free sulphate and the sulphato complex:

$$[SO_4]_{L,t} = [SO_4^{2-}] + [S1SO_4^{+}] = [SO_4^{2-}] \{ 1 + f^{3/4}\ 10^{11.7-3pH} \}$$

Therefore:

$$[SO_4^{2-}] = \frac{[SO_4]_{L,t}}{1 + f^{3/4}\ 10^{11.7-3pH}}$$

and

$$[AlSO_4^{+}] = [SO_4]_{L,t} - [SO_4^{2-}] = \left[\frac{f^{3/4}10^{11.7-3pH}}{1 + f^{3/4}10^{11.7-3pH}}\right] [SO_4]_{L,t}$$

Note that in the ionic balance the $[AlSO_4^{+}]$ (a cation) and the $[SO_4^{2-}]$ (an anion) appear on the left and the right side of the equation.

$$\ldots + [AlSO_4^{+}] = 2[SO_4^{2-}] + \ldots$$

$$\ldots + \left[\frac{f^{3/4}\ 10^{11.7-3pH}}{1 + f^{3/4}\ 10^{11.7-3pH}}\right] = 2 \left[\frac{1}{1 + f^{3/4}\ 10^{11.7-3pH}}\right] + \ldots$$

Since both terms have a common denominator, they may be combined to yield:

$$\ldots = \left[\frac{2 - f^{3/4}\ 10^{11.7-3pH}}{1 + f^{3/4}\ 10^{11.7-3pH}}\right] \ldots$$

Lake pH Model

From the above, the pH equation for the lake becomes:

$$3\left[\frac{10^{8.5-3pH}}{f^{9/4}}\right]+2\left[\frac{10^{3.91-2pH}}{f}\right]-\left(10^{0.3-pH}\right)+\left(10^{-3.86}\right)+2[Ca]_{L,t}=$$

$$[SO_4]_{L,t}\left[\frac{2-f^{3/4}\ 10^{11.7-3pH}}{1+f^{3/4}\ 10^{11.7-3pH}}\right]+\left[\frac{10^{pH-11.0}}{f^{1/4}}\right]$$

where

f = the activity coefficient; for the lake environment f would be expected to fall between 0.6 an 0.9 with a probable value of 0.8.

The above pH equation requires an iterative solution.

REFERENCES

Ontario Ministry of the Environment, 1972 to 1977. "Water Quality Data for Ontario Lakes and Streams - Volumes VII, VIII, IX, X, XI, XII." Water Resources Branch Annual Publication.

Stumm, W. and Morgan, J.J., 1981. "Aquatic Chemistry: An Introduction Emphasizing Chemical Equilibrium in Natural Waters." John Wiley and Sons, New York.

KEY TO TABLE C.1

Distribution Type		Distribution Specifications
C	Constant	single numerical value
LN	Lognormal	geometric mean, geometric standard deviation, minimum, maximum
N	Normal	arithmetic mean, arithmetic standard deviation, minimum, maximum
T	Triangular	minimum, mode (peak), maximum

Table C.1

Lake Model Variables

Symbol	Description	Units	Nominal Value	Probability Distribution	Computer Source	Symbol
P	net precipitation	$m.yr^{-1}$	0.4	N(0.4, 0.1, 0.1, 0.7)	INPUT	NPRCP
$\{Ra\}_{w,t}$	radium activity in lake water at time t	Bq	calculated	calculated	LAKE	RAWAT
$[Ra]_{T,t}$	concentration of radium in the contaminated tailings area runoff at time t	$Bq.m^{-3}$	calculated	calculated	SOURCE	WCONC(7,4,2)
A_T	tailings area	m^2	$5.0x10^5$	$C(5.0x10^5)$	INPUT	ATAIL
$\{Ra\}_{s,t}$	radium activity in exchange zone of sediment at time t	Bq	calculated	calculated	LAKE	RASED
k_1	rate of radium removal from the lake water column to lake sediment	yr^{-1}	0.43	LN(0.43,1.85, 0.13, 1.47)	INPUT	XK1
k_2	rate of radium release from lake sediment to the water column	yr^{-1}	calculated	calculated	LAKE	XK2
k_3	rate of radium removal (burial) to deep lake sediment	yr^{-1}	0.022	LN(0.022,1.2, 0.015, 0.032)	INPUT	XK3
A_W	watershed area	m^2	$2x10^7$	$C(2x10^7)$	INPUT	AWAT
D_L	mean lake depth	m	10	C(10)	INPUT	DLAKE
A_L	mean lake area	m^2	$1.6x10^6$	$C(1.6x10^6)$	INPUT	ALAKE
$[Ra]_{L,t}$	radium concentration in lake water at time t	$Bq.m^{-3}$	calculated	calculated	LAKE	WCONC (7,5,2)
β	parameter relating the rate constant k_2 to rate constant k_1	-	0.18	LN(0.18, 1.35, 0.01 0.33)	INPUT	BETA
T_t	retention time in lake	yr	calculated	calculated	LAKE	1/XLAML
V	lake volume	m^3	$1.6x10^7$	$C(1.6x10^7)$	INPUT	VLAKE

Table C.1 (continued)

Symbol	Description	Units	Nominal Value	Probability Distribution	Computer Source	Symbol
$[Ca]_{in,t}$	calcium concentration in inflow to lake at time t	$mol.L^{-1}$	calculated	calculated	INPUT	CALIN
$[Ca]_{T,t}$	calcium concentration in contaminated tailings area runoff at time t	$mol.L^{-1}$	calculated	calculated	INPUT	WCONC(1,4,2)
$[Ca]_{L,t}$	calcium concentration in the lake at time t	$mol.L^{-1}$	calculated	calculated	INPUT	WCONC(1,5,2)
$[SO_4]_{in,t}$	sulphate concentration in inflow to lake at time t	$mol.L^{-1}$	calculated	calculated	INPUT	SULIN
$[SO_4]_{T,t}$	sulphate concentration in the contaminated tailings area runoff at time t	$mol.L^{-1}$	calculated	calculated	INPUT	WCONC(4,4,2)
$[SO_4]_{L,t}$	sulphate concentration in the lake at time t	$mol.L^{-1}$	calculated	calculated	INPUT	WCONC(4,5,2)
f	activity coefficient for lake pH calculations	-	0.8	T(0.6,0.8,0.9)	INPUT	ACTCO

APPENDIX D

AIR DISPERSION COMPONENT MODEL

APPENDIX D

AIR DISPERSION COMPONENT MODEL

In Phase I (SENES 1984) an atmospheric dispersion model was presented. Because this model was considered adequate for the Phase II assessment it was preserved in its entirety.

The dispersion equation used is the sector-averaged Gaussian plume model. The basic form of the equation is:

$$C_{ai} = \left[\frac{2.032\ f\ Q_i}{u\ \sigma_z\ x} \right] \exp \left[\frac{-H^2}{2\ \sigma_z^{\ 2}} \right]$$

where

$C_{ai} = C_{ai}(t)$ = atmospheric concentration of the ith radionuclide at the receptor at time t ($Bq \cdot m^{-3}$)

f = fraction of the time the wind blows from the source to the receptor in one of 16 sectors

$Q_i = Q_i(t)$ = emission rate of the ith radionuclide at time t ($Bq \cdot s^{-1}$)

u = mean wind speed from the source to the receptor ($m.s^{-1}$)

σ_z = vertical dispersion coefficient (m)

x = downwind distance from the source to the receptor (m)

H = height of the source (m)

This equation strictly applies to a point source. To accommodate an area source the well known concept of a virtual source should be used, as discussed in the Phase I report (SENES 1984). Because virtual distance is dependent on the atmospheric stability scenario and the downwind distance chosen for the receptor, it must be evaluated on a case by case basis. For the Phase II assessment the source/receptor relationship was taken to be the same as for Phase I, that is the virtual distance was set at 3600 m (2000 m + 1600 m virtual setback) as calculated in the Phase I report (SENES 1984).

The meteorological data used in this model were developed from data representative of conditions that occur at a selected location on the Canadian Shield. In particular the Atmospheric Environment Service climatological station at Gore Bay, Ontario was the primary source. These data are identical to those used in the Phase I model (SENES 1984). A summary description of the air dispersion model variables and the attributes assigned to the probability distributions is given on Table D.1.

REFERENCES

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada - Phase I." Prepared for the National Uranium Tailings Program, CANMET, Energy, Mines and Resources Canada.

KEY TO TABLE D.1

Distribution Type		Distribution Specifications
C	Constant	single numerical value
LN	Lognormal	geometric mean, geometric standard deviation, minimum, maximum
N	Normal	arithmetic mean, arithmetic standard deviation, minimum, maximum
U	Uniform	minimum, maximum

Table D.1

Air Dispersion Model Variable

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$C_{ai}(t)$	atmospheric conc. of ith radionuclide at time t at the receptor	$Bq.m^{-3}$	calculated	calculated	AIRDIS	CAIR
f	fraction of time wind blows from the source to the receptor	-	0.1	N(0.1 , 0.02, 0.06 0.14)	INPUT	FWIND2
H	height of the source	m	2	U(1,3)	INPUT	HSRC
$Q_i(t)$	emission rate of ith radionuclide at time t	$Bq.s^{-1}$	calculated	calculated	DUST	EMRATE
u	mean annual wind speed	$m.s^{-1}$	5.3	N(5.3, 0.5, 3.8, 6.8)	INPUT	UWIND
x	downwind distance to receptor	m	3600	C (3600)	INPUT	XDIST
σ_z	vertical dispersion coefficient	m	60	LN (60, 1.08, 47.6, 75.7)	INPUT	VDISP

APPENDIX E

THE CALCULATION OF THE CONCENTRATION OF RADIONUCLIDE i IN ENVIRONMENTAL MEDIA

APPENDIX E

TABLE OF CONTENTS

Page

1.0 TERRESTRIAL MODELS E-1

1.1 Introduction E-1

1.2 Concentration of Radionuclide i in the Edible Portion of Vegetation Due to Direct Deposition from the Air (SUBROUTINE VEGDEP) E-1

1.3 Concentration of Radionuclide i in Animal Produce Due to Vegetation Ingestion (SUBROUTINE VEGEAT) E-3

2.0 AQUATIC MODELS E-4

2.1 Concentration of Radionuclide i in Edible Portions of Aquatic Animals (SUBROUTINE FISH) E-4

REFERENCES E-6

THE CALCULATION OF THE CONCENTRATION OF RADIONUCLIDE i IN ENVIRONMENTAL MEDIA

1.0 TERRESTRIAL MODELS

1.1 Introduction

Appendices C and D describe how the concentration of a radionuclide in surface water and air were modelled, respectively. This appendix briefly describes how the concentration of a radionuclide was calculated in other environmental media, for use in subsequent calculations of the dose received by the critical receptor (see Appendix F).

No major changes were made to the models used in Phase I (SENES 1984). However, as discussed in the main body of the report, regression analysis is considered to be a useful tool for sensitivity and uncertainty analysis of predicted values. Therefore, it was considered desirable, to vary some parameters previously treated as nominal constants (SENES 1984) to permit their inclusion in a regression analysis. The extent of this variation was based on the collective opinion of the study group as to how certain they were about the nominal value of particular parameters. Such parameters were typically allowed to vary 10 to 50% about the nominal values and were given a triangular or uniform distribution.

1.2 Concentration of Radionuclide i in the Edible Portion of Vegetation Due to Direct Deposition from the Air (Subroutine VEGDEP)

Airborne materials can be transferred to the edible portions of vegetables consumed by man or to forage consumed by animals either by direct foliar retention or root uptake. Subroutine

VEGDEP takes into account only direct foliar retention. Root uptake was not included in the model because of the relatively small contribution it makes to the total vegetation concentration. (See Appendix J, in Volume 2).

Four types of vegetation were modelled; lichens (winter forage for caribou), grass (summer forage for caribou), above-ground vegetables, and below-ground vegetables. In the following model description those parameters that may vary with the type of vegetation are subscripted with a v. Those parameters that may vary with the radionuclide are subscripted with an i, although only radium was modelled in this phase of model development.

The following equation was used to calculate the average annual concentration of radionuclide i in vegetation type v (U.S. NRC 1982):

$$C_{vi} = C_{ai} \; V_d \; F_{rv} \; E_v \; \frac{1 - \exp(-\lambda_{wv} t_v)}{Y_v \; \lambda_{wv}} \qquad (1.1)$$

where

C_{vi} = $C_{vi}(t)$ is the resulting concentration of radionuclide i in vegetation v at time t ($Bq.g^{-1}$ (wet weight))

C_{ai} = $C_{ai}(t)$ is the average annual concentration of radionuclide i in the air at that location at time t($Bq.m^{-3}$)

V_d = the deposition velocity ($m.s^{-1}$)

F_{rv} = the fraction of the foliar deposition retained on edible portions of vegetation type v (0 to 1)

E_v = the fraction of the total deposition retained on the plant surface of vegetation type v (0 to 1)

λ_{wv} = a decay constant accounting for weathering losses from vegetation type v (s^{-1})

t_v = the duration of the plant exposure to atmospheric deposition for vegetation type v (s)

Y_v = the yield density for vegetation type v (g(wet weight).m^{-2})

Note that time dependence is due only to the time dependence of the air concentration. These calculations were not done by particle size, rather a typical particle size was assumed.

Table E.1 lists the variables used in this model. The values for F_{rv}, E_v, t_v, Y_v and λ_{wv} for vegetables (above and below ground) and pasture grass were allowed to vary uniformly around the nominal value recommended by the U.S. NRC (1982), with one exception. F_{rv} for pasture grass was taken from the discussion paper by Hoffman et al in Volume 2.

Lichens are not included in the U.S. NRC recommendations. All nominal values were based on previous investigations (SENES 1983) and allowed to vary around the nominal values as above. Briefly, the high values for F_{rv} and E_v reflect the high surface area and covering ability (in terms of the ground surface) of lichens, and the fact that basically all of the plant is edible. The values for t_v and λ_{wv} reflect the long life span of lichens (basically limited only by grazing by caribou) and high retention of any debris falling on them.

1.3 Concentration of Radionuclide i in Animal Produce Due to Vegetation Ingestion (Subroutine VEGEAT)

Radionuclides deposited on winter (lichen) and summer (grass) forage may be eaten by caribou which in turn may be eaten by man. The transfer of radionuclide i from vegetation to meat was modelled as follows (U.S. NRC 1982):

$$C_{bi} = Q_b F_{bi} \sum_{v} F_v \, C_{vi} \tag{1.2}$$

where

$C_{bi} = C_{bi}(t)$ is the resulting annual average concentration of radionuclide i in caribou meat at time t ($Bq.g^{-1}$ (wet weight))

Q_b = the average annual feed ingestion rate of the caribou (g (wet weight).yr^{-1})

F_{bi} = the feed-to-meat transfer coefficient for caribou ($Bq.g^{-1}$ (wet wt). $Bq^{-1}.yr$)

F_v = the fraction of the annual intake that is local vegetation type v (lichen and grass in this case) (0 to 1)

C_{vi} = $C_{vi}(t)$ is the concentration of radionuclide i in vegetation type v (lichen and grass in this case) at time t ($Bq.g^{-1}$ (wet weight))

The time dependence in this case is only due to the time dependence in the vegetation concentration. Table E.2 summarizes these variables and their distributions.

Q_b and F_{bi} actually represent the feed ingestion rate and feed-to-meat transfer coefficient for cattle (U.S.NRC 1982), in the absence of caribou-specific information.

The fraction of the annual intake of the caribou that were local lichen and grass were allowed to vary independently between 0.05 and 0.5. Thus the total local annual feed intake will seldom add up to 100 per cent. This was included to try to reflect the migratory nature of caribou, as it was considered too conservative to assume 100 per cent consumption of local feed.

2.0 AQUATIC MODELS

2.1 Concentration of Radionuclide i in Edible Portions of Aquatic Animals (Subroutine FISH)

A simple aggregate model was used to calculate the concentration of radionuclide i in the flesh of fish from the lake (U.S. NRC 1982):

$$C_{fi} = B_{fi} C_{swi}$$

where

C_{fi} = $C_{fi}(t)$ is the resulting annual average concentration of radionculide i in fish flesh at time t ($Bq.g^{-1}$ (wet weight))

B_{fi} = is water to fish-flesh transfer coefficient for radionuclide i ($Bq.g^{-1}$ (wet weight)$.Bq^{-1}.m^{3}$)

C_{swi} = $C_{swi}(t)$ is the calculated average annual concentration of radionuclide i in the lake water at time t ($Bq.m^{-3}$)

The distribution for B_{fi} was based on a review by SENES (Appendix H, Volume 2), the results of which are summarized in Table E.3.

REFERENCES

SENES Consultants Limited, 1983. "Decommissioning of the Beaverlodge Mine/Mill Operations and Reclamation of the Site: Environmental and Radiological Pathways Analysis." Prepared for Eldorado Resource Limited.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada - Phase I." Prepared for the National Uranium Tailings Program, CANMET, Energy Mines and Resources Canada.

United States Nuclear Regulatory Commission, 1982. "Calculational Models for Estimating Radiation Doses to Man from Airborne Radioactive Materials Resulting from Uranium Milling Operations." Regulatory Guide 3.51, Office of Nuclear Regulatory Research.

KEY TO TABLES E.1, E.2 and E.3

Distribution Type		Distribution Specifications
C	Constant	single numerical value
LN	Lognormal	geometric mean, geometric standard deviation, minimum, maximum
N	Normal	arithmetic mean, arithmetic standard deviation, minimum, maximum
T	Triangular	minimum, mode (peak), maximum
U	Uniform	minimum, maximum
LU	Log Uniform	minimum, maximum

Table E.1

Direct Deposition to Vegetation Model Variables

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$C_{ai}(t)$	air concentration of radionuclide i	$Bq.m^{-3}$	calculated	calculated	AIR	CAIR
$C_{vi}(t)$	vegetation type v concentration radionuclide i:	$Bq.g^{-1}$	calculated	calculated	VEGDEP	CVEG(J)
						J=vegetation type
	lichens					CVEG(1)
	pasture grass					CVEG(2)
	above ground veg.					CVEG(3)
	below ground veg.					CVEG(4)
F_{rv}	fraction of total deposition retained on surface of vegetation:	-			INPUT	FDEPSF(J
	lichens		0.95	U(0.8, 1.0)		FDEPSF(1)
	pasture grass		0.2	N(0.47, 0.3, 0.01, 1.0)		FDEPSF(2)
	above ground veg.		0.2	U(0.1, 0.3)		FDEPSF(3)
	below ground veg.		0.2	U(0.1, 0.3)		FDEPSF(4)
E_v	fraction of plant deposition retained on edible portions:	-			INPUT	FDEPED(J)
	lichens		1.0	U(0.8, 1.0)		FDEPED(1)
	pasture grass		1.0	U(0.8, 1.0)		FDEPED(2)
	above ground veg.		1.0	U(0.8, 1.0)		FDEPED(3)
	below ground veg.		0.1	U(0.05, 0.15)		FDEPED(4)
t_v	duration of exposure to deposition:	s			INPUT	DUREXP(J)
	lichens		100×10^6	$U(90x10^6, 110x10^6)$		DUREXP(1)
	pasture grass		2.952×10^6	$U(2.7x10^6, 3.1x10^6)$		DUREXP(2)
	above ground veg.		5.184×10^6	$U(4.7x10^6, 5.6x10^6)$		DUREXP(3)
	below ground veg.		5.184×10^6	$U(4.7x10^6, 5.6x10^6)$		DUREXP(4)
V_d	deposition velocity	$m.s^{-1}$	0.02	U(0.003, 0.03)	INPUT	DEPVEL
Y_v	yield density:	$g.m^{-2}$			INPUT	YDEN(J)
	lichens		500	U(400,600)		YDEN(1)
	Pasture grass		750	U(500,1000)		YDEN(2)
	above ground veg.		2000	U(1500,2500)		YDEN(3)
	below ground veg.		2000	U(1500,2500)		YDEN(4)
λ_{wv}	decay constant accounting for weathering loss:	s^{-1}			INPUT	VEGLAM(J)
	lichens		5.73×10^{-7}	$U(2x10^{-9}, 4x10^{-9})$		VEGLAM(1)
	pasture grass		5.73×10^{-7}	$U(4x10^{-7}, 6x10^{-7})$		VEGLAM(2)
	above ground veg.		5.73×10^{-7}	$U(4x10^{-7}, 6x10^{-7})$		VEGLAM(3)
	below ground veg.		5.73×10^{-7}	$U(4x10^{-7}, 6x10^{-7})$		VEGLAM(4)

Table E.2

Vegetation to Animal Produce Model Variables

Symbol	Description	Units	Nominal	Probabilistic	Source	Computer Symbol
$C_{bi}(t)$	concentration of radionuclide i in in animal produce: caribou meat	$Bq.g^{-1}$ (wet wt)	calculated	calculated	VEGEAT	CMEAT
$C_{vi}(t)$	vegetation type v concentration radionuclide i:	$Bq.g^{-1}$ (wet wt)	calculated	calculated	VEGDEP	CVEG(J)
	lichen					CVEG(1)
	pasture grass					CVEG(2)
F_{bi}	feed-to-meat transfer coefficient (Ra-226): caribou	$Bq.g^{-1}$ (wet wt). $Bq^{-1}.yr$	1.4×10^{-9}	$U(0.3 \times 10^{-9}, 2.5 \times 10^{-9})$	INPUT	FBI
F_v	fraction of annual intake that is local vegetation:	-			INPUT	FAUT(J)
	lichen		0.6	T (0.05, 0.2, 0.5)		FAUT(1)
	pasture grass		0.4	T (0.05, 0.2, 0.5)		FAUT(2)
Q_b	feed ingestion rate: caribou	$g(wet\ wt).yr^{-1}$	2555×10^3	$C\ (2500 \times 10^3)$	INPUT	FIR

Table E.3

Water to Aquatic Produce Model Variables

Symbol	Description	Units	Nominal	Probabalistic	Source	Computer Symbol
B_{fi}	water to aquatic produce transfer coefficient (Ra-226) fish flesh	$Bq.g^{-1}$ (wet wt) $Bq^{-1}.m^3$	1.5×10^{-6}	LN(16×10^{-6}, 3.5, 1.0×10^{-6}, 600×10^{-6})	INPUT	BF
$C_{fi}(t)$	concentration of radionuclide i: fish flesh	$Bq.g^{-1}$	calculated	calculated	FISH	CFISH
$C_{swi}(t)$	concentration of radionuclide i in lake water	$Bq.m^{-1}$	calculated	calculated	LAKE	WCONC(7,5,2)

APPENDIX F

THE CALCULATION OF THE
ANNUAL DOSE RECEIVED
BY THE CRITICAL RECEPTOR

APPENDIX F

TABLE OF CONTENTS

Page

1.0 INTRODUCTION F-1

2.0 INGESTION DOSE
  2.1 Annual Dose Due to the Ingestion of Vegetation F-1
  2.2 Annaul Dose Due to the Ingestion of Animal Produce F-3
  2.3 Annual Dose Due to the Ingestion of Drinking Water F-3
  2.4 Annual Dose Due to the Ingestion of Fish F-4

3.0 INHALATION DOSE F-5
  3.1 Annual Dose Due to the Inhalation of Suspended Particulates F-5
  3.2 Annual Dose Due to the Inhalation of Radon and Its Daughters F-6

REFERENCES

APPENDIX F
THE CALCULATION OF THE ANNUAL DOSE RECEIVED BY THE CRITICAL RECEPTOR

1.0 INTRODUCTION

Due to the similarity of the models detailed below, they were grouped into one subroutine (DOSEIN). The variables for all are listed in Table F.1.

As discussed in the main body of this report, regression analysis is considered to be a most useful tool for sensitivity and uncertainty analysis of predicted values. It was considered desirable, therefore, to vary some parameters (such as fraction of intake that is local) previously treated as constants (SENES 1984) to permit their inclusion in regression analysis. The extent of this variation was based on the opinion of the study group as to how certain they were about the nominal value of the particular parameter. Such parameters were typically allowed to vary 10 to 50% about the nominal values and were given triangular or uniform distributions. Classes of parameters specifically excluded from this regression analysis included the annual intakes of food (Nutrition Canada n.d.), water, and air (ICRP 1975) and dosimetry (Johnson and Dunford 1983).

2.0 INGESTION DOSE

2.1 Annual Dose Due to the Ingestion of Vegetation

The annual average effective whole-body dose equivalent commitment (herein called annual average dose) received by an individual due to the ingestion of above- and below-ground vegetables was calculated as follows:

$$d_{vi}(ing) = DCF_i(ing) \sum_v U_v\, C_{vi} R_v L_v \qquad (2.1)$$

where

$d_{vi}(ing)$ = $d_{vi}(ing)(t)$ is the average annual whole body dose due to the ingestion of radionuclide i in vegetables v (in this case above and below ground vegetables) at time t ($Sv.yr^{-1}$)

$DCF_i(ing)$ = the whole body dose conversion factor for the ingestion of radionuclide i ($Sv.Bq^{-1}$)

U_v = the average annual intake of vegetable type v (in this case above and below ground vegetables) (g (wet wt)$.yr^{-1}$)

C_{vi} = $C_{vi}(t)$ is the calculated average annual concentration of radionuclide i in vegetation type v at time t ($Bq.g^{-1}$ (wet weight))

R_v = the fraction of activity in vegetable type v remaining after food preparation (0 to 1)

L_v = the fraction of the vegetable type v that is grown locally (0 to 1)

Note that the time dependence in $d_{vi}(ing)(t)$ is due only to the time dependence of $C_{vi}(t)$.

R_v was chosen to be uniformly distributed from 0.5 to 1.0. The lower value is that recommended by the U.S. NRC (1982). The large upper value was added in recognition of the uncertainty of this number.

Lv was assigned a triangular distribution from 0.1 to 0.6, with a peak at 0.3. The range, although somewhat arbitrary, was chosen in recognition of the fact that it would be very unlikely that someone could or would grow a large percentage of their vegetables in the area being modelled (i.e. the Canadian Shield).

2.2 Annual Dose Due to the Ingestion of Animal Produce

The model used for calculating the average annual dose received by an individual due to the ingestion of caribou meat was (U.S. NRC 1982):

$$d_{bi}(ing) = DCF_i(ing)\ U_b\ C_{bi}\ L_b \qquad (2.2)$$

where

$d_{bi}(ing)$ = $d_{bi}(ing)(t)$ is the average annual whole body dose due to the ingestion of radionuclide i in caribou meat at time t ($Sv.yr^{-1}$)

$DCF_i(ing)$ = the whole body dose conversion factor for the ingestion of radionuclide i ($Sv.Bq^{-1}$)

Ub = the average annual meat intake for an Ontario male, age 20 to 39 (Nutrition Canada n.d.), assumed to be all caribou ($g.yr^{-1}$)

C_{bi} = $C_{bi}(t)$ is the calculated average annual concentration of radionuclide i in caribou meat at time t ($Bq.g^{-1}$)

L_b = the fraction of the annual intake that is local caribou meat (0 to 1)

The fraction of the annual intake that is local (i.e. from a caribou killed in the immediate area) was allowed to range from 0 to 100% as this would vary greatly from year to year.

2.3 Annual Dose Due to the Ingestion of Drinking Water

The average annual dose received by an individual as a result of drinking water from the lake was modelled as (U.S. NRC 1982):

$$d_{wi}(ing) = DCF_i(ing)\ U_w\ C_{swi}\ L_w \qquad (2.3)$$

where

$d_{wi}(ing)$ = $d_{wi}(ing)(t)$ is the average annual whole body dose due to the ingestion of radionuclide i in drinking water at time t ($Sv.yr^{-1}$)

$DCF_i(ing)$ = the whole body dose conversion factor for the ingestion of radionuclide i ($Sv.Bq^{-1}$)

U_w = the average annual intake of drinking water ($m^3 \cdot yr^{-1}$)

C_{swi} = $C_{swi}(t)$ is the calculated average annual concentration of radionuclide i in the lake water at time t ($Bq.m^{-3}$)

L_w = the fraction of the annual intake that is local drinking water (0 to 1)

The fraction of the annual water consumption that is from the lake was set equal to the fraction of the year that the critical receptor stays at the site (see Section 3.1 of this Appendix). The annual water consumption is for Reference Man (ICRP 1975).

2.4 Annual Dose Due to the Ingestion of Fish

The average annual dose received by an individual as a result of eating fish from the lake was modelled as follows (U.S. NRC 1982):

$$d_{fi}(ing) = DCF_i(ing)\ U_f\ C_{fi}\ L_f \qquad (2.4)$$

where

$d_{fi}(ing)$ = $d_{fi}(ing)(t)$ is the average annual whole body dose due to the ingestion of radionculde i in fish at time t ($Sv.yr^{-1}$)

$DCF_i(ing)$ = the whole body dose conversion factor for the ingestion of radioncluide i ($Sv.Bq^{-1}$)

U_f = the average annual intake of fish ($g.yr^{-1}$)

C_{fi} = $C_{fi}(t)$ is the calculated average annual concentration of radionuclide i in fish at time t ($Bq.g^{-1}$)

L_f = the fraction of the annual intake that is local fish (0 to 1)

The average intake of fish is for an average Ontario male (age 20 to 39) (Nutrition Canada n.d.). It was assumed that all of his yearly intake of fish is from this lake as it would be the easiest source for this individual.

3.0 INHALATION

3.1 Annual Dose Due to the Inhalation of Suspended Particulates

The average annual dose received by an individual due to inhalation (excluding radon) was calculated as follows:

$$d_i(inh) = DCF_i(inh)\ U_a\ C_{ai}\ OF \qquad (3.1)$$

where

$d_i(inh)$ = $d_i(inh)(t)$ is the average annual whole body dose due to the inhalation of radionuclide i at time t ($Sv.yr^{-1}$)

$DCF_i(inh)$ = the whole body dose conversion factor for the inhalation of radionuclide i ($Sv.Bq^{-1}$)

U_a = the average annual inhalation rate of the individual ($m^3.yr^{-1}$)

C_{ai} = $C_{ai}(t)$ is the calculated average annual air concentration of radionuclide i at time t ($Bq.m^{-3}$)

OF = the fraction of the year that is spent at that location

The annual inhalation rate was taken from Reference Manual (ICRP 1975). The occupancy factor was chosen to have a nominal value of 0.6 based on four months away for hunting and trapping, plus a few trips for obtaining supplies, visiting, etc. This was allowed to vary from 0.4 to 0.9 (Table F.1) for the purposes of regression analysis.

3.2 Annual Dose Due to the Inhalation of Radon and Its Daughters

The radon source and air dispersion models result in an estimation of the radon-in-air concentration at the receptor of interest. This concentration must be converted to a dose (actually a dose rate measured in units of $Sv.yr^{-1}$) so that it can be added to the doses from the other exposure pathways being evaluated.

This conversion requires an estimation of the equilibrium factor, F, between radon and radon daughters. Theoretically, F can range from 0.0 (no radon daughters) to 1.0 (complete radon-daughter ingrowth). The value of F at the receptor depends only on the transit time between the radon source and the receptor, if it is assuming that there are no radon daughter removal mechanisms other than radioactive decay. Since in reality the daughters are removed by several other mechanisms that are difficult to quantify (e.g. attachment to dust that falls to the ground) and since the major exposure occurs indoors where F is dependent on several additional factors (such as ventilation rate), it has been customary to select, (rather than calculate), an F factor that is representative of annual average indoor conditions. (The radon daughter ingrowth outdoors is negligible because the source-receptor transit time is typically short, less than 5 minutes). In the Phase I study, an F value of 0.5 was assumed to be typical. For the Phase II study, the assumed F value is replaced by an input distribution, triangular in shape, with a range of 0.2 to 0.7 and a mode of 0.4.

Therefore, based on an assumed equivalence of 4.8 WLM and 0.05 Sv (ICRP, 1981), a fractional 0.75 of the time spent indoors, and the definitions of working level (WL, the concentration of radon daughters equivalent to 100 $pCi.L^{-1}$ of radon in equilibrium), working level month, (WLM, exposure to 1 WL for 170 h), and occupancy factor (OF, the fraction of time at the receptor site, an input variable), the conversion between exposure to 1 $Bq.m^{-3}$ of radon and annual dose is given by:

$$\begin{aligned} & 0.75 \times OF \times 1\ Bq.m^{-3} \times 10^{-3}\ m^{3}.L^{-1} \times 27\ pCi.Bq^{-1} \\ & \quad \times F\ WL/(100\ pCi.L^{-1}) \times 8760\ h.yr^{-1} \times WLM/(WL \times 170\ h) \\ & \quad \times 0.05\ Sv/4.8\ WLM \\ & \quad = OF \times F \times 1.09 \times 10^{-4}\ Sv.yr^{-1} \qquad (3.2) \end{aligned}$$

(It should be noted that the (risk) equivalence of 4.8 WLM and 0.05 Sv is considered debatable by some health physicists but examination of this equivalence is beyond the scope of this report. See Stewart (1984) for a detailed analysis). When F=0.5 and indoor occupancy at the site is 100%, instead of 75%, this reverts to the Phase I dose conversion factor of OF x 7.2×10^{-5} $Sv.yr^{-1}$ per $Bq.m^{-3}$.

REFERENCES

International Commission on Radiological Protection,(ICRP), 1975. "Report of the Task Group on Reference Man." ICRP Publication 23, Pergamon Press.

International Commission on Radiological Protection (ICRP), 1981. "Limits for Inhalation of Radon Daughters by Workers." ICRP Publication 32, Annals of the ICRP, Vol. 6, No. 1.

Johnson, J.R and Dunford, D.W., 1983. "Dose Conversion Factors for Intakes of Selected Radionuclides by Infants and Adults." Atomic Energy of Canada Limited Report AECL-7919, Chalk River Nuclear Laboratories, Chalk River, Ontario.

Nutrition Canada, not dated. "Food Consumption Patterns Report." Prepared by the Department of National Health and Welfare.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada. Phase I." Prepared for the National Uranium Tailings Program, CANMET, Energy Mines and Resources Canada.

Stewart, C.G., 1984. "Comments Upon 'Basis of Intake Limits for Radon and Thoron Daughters' Proposed by the Atomic Energy Control Board of Canada in Consultative Document C-78 Entitled 'Limitation of Exposure to Ionizing Radiation' Issued for Comment 14 November 1983." Prepared in Association with SENES Consultants Limited for the Canadian Nuclear Association.

United States Nuclear Regulatory Commission, (NRC) 1982. "Calculational Models for Estimating Radiation Doses to Man from Airborne Radioactive Materials Resulting from Uranium Milling Operations." Regulatory Guide 3.51, Office of Nuclear Regulatory Research.

KEY TO TABLE F.1

Distribution Type		Distribution Specifications
C	Constant	single numerical value
LN	Lognormal	geometric mean, geometric standard deviation, minimum, maximum
N	Normal	arithmetic mean, arithmetic standard deviation, minimum, maximum
T	Triangular	minimum, mode (peak), maximum
U	Uniform	minimum, maximum
LU	Log Uniform	minimum, maximum

Table F.1

Variables Used in the Calculation of Annual Dose to the Critical Receptor

Symbol	Description	Units	Nominal	Probabilistic	Computer Source	Symbol
Vegetation Model						
d_{vi}(ing)(t)	annual whole body dose commitment due to vegetable v ingestion for ith radionuclide	$Sv \cdot yr^{-1}$	calculated	calculated	DOSEIN	VEGDOS
DCF_i(ing)	whole body DCF for for ingestion of radionuclide i:Ra-226	$Sv \cdot Bq^{-1}$	3.2×10^{-7}	C (3.2×10^{-7})	INPUT	DCFING
U_v	average annual intake of vegetable type v:	g(wet wt)$.yr^{-1}$			INPUT	
	above ground veg.		157×10^3	C (157×10^3)		VEGINA
	below ground veg.		70×10^3	C (70×10^3)		VEGINB
C_{vi}(t)	concentration of radionuclide i in vegetation type v:	$Bq \cdot g^{-1}$	calculated	calculated	VEGDEP	CVEG(J)
	above ground veg.					CVEG(3)
	below ground veg.					CVEG(4)
R_v	fraction of activity in vegetable type v after food preparation:				INPUT	
	above ground veg.		1.0	U(0.5, 1.0)		FACTA
	below ground veg.		1.0	U(0.5, 1.0)		FACTB
L_v	fraction of vegetable type v grown locally:				INPUT	
	above ground veg.		0.5	T(0.1,0.3,0.6)		FLOCA
	below ground veg.		0.5	FLOCA		FLOCB

Table F.1 continued

Symbol	Description	Units	Nominal	Probablistic	Source	Computer Symbol
Fish Ingestion Model						
$d_{fi}(ing)(t)$	whole body dose commitment due to ingestion of radionuclide i in fish	$Sv.yr^{-1}$	calculated	calculated	DOSEIN	FISHDS
U_f	average annual intake of fish	$g.yr^{-1}$	8×10^3	$U(4x10^3, 12x10^3)$	INPUT	FISHIN
$C_{fi}(t)$	annual concentration of radionuclide i in fish	$g.yr^{-1}$	calculated	calculated	FISH	CFISH
L_f	fraction of annual intake that is local fish		1.0	1.0	INPUT	FFISH
Inhalation Model (Particulates)						
$d_i(inh)(t)$	whole body dose commitment due to inhalation of radionuclide i	$Sv.yr^{-1}$	calculated	calculated	DOSEIN	XINDOS
$DCF_i(inh)$	whole body DCF due to inhalation of radionuclide i: Ra-226	$Sv.Bq^{-1}$	2.6×10^{-6}	$C\ (2.6 \times 10^{-6})$	INPUT	DCFINH
U_a	annual inhalation rate	$m^3.yr^{-1}$	8400	C (8400)	INPUT	XINHIN
$C_{ai}(t)$	average annual air concentration of radionuclide i	$Bq.m^{-3}$	calculated	calculated	AIRDIS	CAIR
OF	fraction of year in location		0.6	T (0.4,0.6,0.9)	INPUT	OF
Inhalation Model (Radon)						
$d_{Rn}(inh)(t)$	whole body dose commitment due to inhalation of radon plus daughters	$Sv.yr^{-1}$	calculated	calculated	DOSEIN	RNDOS
F	equilibrium factor for radon daughters to radon		0.5	T (0.2,0.4,0.7)	INPUT	F

Table F.1 continued

Symbol	Description	Units	Nominal	Probabilistic	Computer Source	Symbol
Caribou Ingestion Model						
$d_{bi}(ing)(t)$	annual whole body dose commitment due to ingestion of radionuclide i in caribou meat	$Sv \cdot yr^{-1}$	calculated	calculated	DOSEIN	XMEATD
U_b	annual intake of caribou meat	$g \cdot yr^{-1}$	79×10^3	C (79×10^3)	INPUT	XMEATI
$C_{bi}(t)$	annual concentration of radionuclide i in caribou meat	$Bq \cdot g^{-1}$	calculated	calculated	VEGEAT	CMEAT
L_b	fraction of annual caribou meat intake that is local	-	0.5	U(0.0, 1.0)	INPUT	FMEAT
Drinking Water Model						
$d_{wi}(ing)(t)$	whole body dose commitment due to ingestion of radionuclide i in drinking water	$Sv \cdot yr^{-1}$	calculated	calculated	DOSEIN	WATDOS
U_w	average annual intake of water	$m^3 \cdot yr^{-1}$	0.6	C (0.6)	INPUT	WATIN
$C_{swi}(t)$	annual concentration of radionuclide i in lake water	$Bq.m^{-3}$	calculated	calculated	LAKE	WCONC(7,5,2)
L_w	fraction of annual water consumption that is local		1.0	OF	INPUT	FWAT

APPENDIX G

GROUNDWATER TRANSPORT COMPONENT MODEL

prepared by
Golder Associates
for
SENES Consultants Limited

APPENDIX G

TABLE OF CONTENTS

Page

1.0 INTRODUCTION 1

2.0 MODELLING APPROACH 4

3.0 MODELLING TIME VARIANT INPUT 9

4.0 MODELLING DEMONSTRATION 11

- 4.1 Simple Source Functions 11
- 4.2 Hypothetical Source Models 12
- 4.3 The Effect of Discretization 14

5.0 REVIEW OF PARAMETERS AND DISTRIBUTIONS 16

6.0 SUMMARY 23

REFERENCES

LIST OF TABLES

1 Input Parameters and Values Used in Modelling Exercises

2 Summary of Parameter Distributions

LIST OF FIGURES

1 Groundwater Transport Component Model Configuration, Input and Output Parameters

2 Handling Time-Variant Input

3 Breakthrough Curve for a Continuous Source

4 Breakthrough Curve for Decreasing Pulse

5 Breakthrough Curve for Increasing Pulse

6 Breakthrough Curve for Increasing/Decreasing Pulse

7 Breakthrough Curve for Decreasing/Increasing Pulse

8 Linear Model

9 Mixed Model

10 Breakthrough Curve for Hypothetical Source Models

11 Effect of Discretization

1. INTRODUCTION

The conventional modelling approach for simulating mass transport in groundwater systems has historically been deterministic in nature. Single value parameters are used to solve a mathematical relationship describing transport by advective and dispersive processes. It is generally accepted that single-number representations of hydrogeologic parameters are not necessarily representative of reality due to the inherent non-uniformity or heterogeneity of porous media on normal scales of observation. This has been pointed out in the literature and there is some question as to the validity of selecting a single value for each flow parameter that is representative, in order to define an 'equivalent' uniform porous medium (Freeze, 1975).

As a result of these difficulties there is a realization that, even in hydrogeological environments where a great deal of data are available, the exact spatial variation of the heterogeneous components remains unknown. The probabilistic approach to groundwater modelling, in which the input parameters are random with an associated probability density, leads to a probability distribution of the output variables which reflects the uncertainty of the model prediction due to heterogeneity of the porous media (Smith and Freeze, 1979). This presentation of results is particularly useful for risk assessment studies.

The task of Golder Associates has been to develop a probabilistic component model for the simulation of the transport of contaminants from a uranium tailings impoundment to a surface water receptor. The component model incorporates simplified one-dimensional steady-state groundwater flow conditions with contaminant-specific retardation parameters to simulate the movement of contaminants through the groundwater system. The model considers the flow path to be isotropic and homogeneous (with parameter values selected at random) and chemical reactions governed by the limitations of the application of distribution coefficients.

An attempt has been made to describe the complex geochemical processes which control radionuclide migration by utilizing a "bulk" retardation factor. The geochemical reactions which will occur in the aquifer and affect ^{226}Ra movement include acid neutralization (dissolution of carbonate minerals), adsorption and coprecipitation. Mobility of radium will initially be controlled by the migration of acidic plumes, which has been observed to be less than a few per cent of the groundwater velocity. At later times ^{226}Ra can be adequately modelled using distribution coefficients for acidic aquifers. The modelling and geochemical equilibria calculations required to fully define these geochemical processes is beyond the scope of the present study. The use of retardation factors provides a conservative analysis which incorporates the many processes involved.

The approach has been to solve the differential equation describing the movement of a radioactive tracer at a specified distance and time. No attempt has been made to track the movement of individual particles as has been done in some random-walk modelling exercises.

The development of the model has evolved through two phases: i) demonstration of the application of the probabilistic approach for groundwater transport; and ii) demonstration of probabilistic transport modelling with a time-variant source input. The key output of the model is the contaminant concentration at the discharge point to the surface water receptor.

This study has been limited in scope to demonstrate the principles of the probabilistic model for a theoretical reference site. The values of controlling parameters have been selected to represent typical uranium tailings environments as described in the literature.

2. MODELLING APPROACH

In this simplified model, it is assumed that the volumetric rate of porewater seepage from the tailings into the groundwater system is equal to the volumetric rate of groundwater discharge to the surface water receptor in the zone influenced by the tailings impoundment. Contaminant concentrations in the groundwater are reduced by adsorption on the geologic materials and by radioactive decay. Consequently, the influx of contaminants to the surface water (Influx = Volumetric Flow X Concentration) will be reduced by these processes. Large-scale mixing or dilution of the tailings porewater seepage with the groundwater is not accounted for in the model. Although this process will reduce contaminant concentrations, volumetric flow rates are correspondingly increased so that this process is of no consequence when contaminant flux rates are calculated.

The model first calculates the groundwater velocity based on:

$$q = KI \qquad (1)$$

where: q is the Darcy velocity (L/T)
K is the hydraulic conductivity (L/T)
I is the hydraulic gradient (L/L)

Input parameters for the longitudinal dispersion, adsorption and radioactive decay of the contaminants are calculated based on:

$$D_\ell = \frac{\alpha q}{n} \qquad (2)$$

where: D_ℓ is the longitudinal dispersion coefficient (L^2/T)
α is the dispersivity (L)
q is the Darcy velocity (L/T)
n is the porosity (L^3/L^3)

$$R = 1 + \frac{\rho b}{n} Kd \tag{3}$$

where: R is the retardation factor

ρb is the bulk density (M/L^3)

n is the porosity (L^3/L^3)

Kd is the distribution coefficient for the contaminant (L^3/M)

$$\lambda = \frac{\ln 2}{t_{\frac{1}{2}}} \tag{4}$$

where: λ is the radioactive decay constant (T^{-1})

$t_{\frac{1}{2}}$ is the radioactive half-life (T)

The movement of a radioactive tracer in a semi-infinite column undergoing continuous radioactive decay can be described by the partial differential equation:

$$\frac{\partial C}{\partial t} = D_\ell \frac{\partial^2 C}{\partial X^2} - \frac{q}{n} \frac{\partial C}{\partial X} - \lambda C \tag{5}$$

The equation can be modified to include chemical retardation where adsorption can be described by an equilibrium isotherm and the reactions are considered to be essentially instantaneous and reversible by including the retardation factor, R;

$$\frac{\partial C}{\partial t} = \frac{D_\ell}{R} \frac{\partial^2 C}{\partial X^2} - \frac{q}{nR} \frac{\partial C}{\partial X} - \lambda C \tag{6}$$

At this time, dispersion due to molecular diffusion has not been included as its effects would be negligible for the advection-dominated environment considered for this particular study.

The contaminant concentration at a distance (X) from the tailings impoundment is calculated for steady-state conditions by:

$$C(X) = C_o \exp\left[\frac{qX}{2nD_\ell}\left[1 - \left[1 + \frac{4\lambda(D_\ell R)n^2}{q^2}\right]^{0.5}\right]\right] \qquad (7)$$

From Bear (1979) p 270

where:

$C(X)$	is the concentration at a distance X (M/L^3)
C_o	is the input concentration (M/L^3)
q	is the Darcy velocity (L/T)
X	is the distance from the tailings (L)
R	is the retardation factor
n	is the porosity (L^3/L^3)
D_ℓ	is the longitudinal dispersion coefficient (L^2/T)
λ	is the radioactive decay constant (T^{-1})

The contaminant flux to the surface water at a distance (X) from the tailings is calculated by:

$$J = C(X)Q \qquad (8)$$

where:

J	is the contaminant flux (M/T)
$C(X)$	is the concentration at distance X (M/L^3)
Q	is the volumetric seepage rate (L^3/T)

In the case of long-term steady-state conditions, contaminant flux to the surface water will only be significantly reduced when the flow path lengths are sufficiently long or the groundwater velocities sufficiently slow to result in long travel times relative to the half-lives of the radioactive contaminants.

The influence of a time variant source input on the discharge concentration will be dependent to a certain extent on the source function itself, and the physical and chemical properties of the flow system and contaminants. For example, a decreasing source concentration will obviously decrease the discharge concentration however, a sufficiently high distribution coefficient will cause a dampening effect for fluctuating source concentrations.

The simulation of the contaminant concentrations in the groundwater under steady-state flow conditions can be provided by:

$$C(X,t) = \frac{C_o}{2} \exp\left[\frac{qX}{2nD_\ell}\right] \bullet$$

$$\left[\exp(-XB) \cdot \operatorname{erfc}\left[\frac{X-(q/nR)^2 + (4\lambda D_\ell/R)^{0.5}t}{2(D_\ell t/R)^{0.5}}\right]\right.$$

$$\left. + \exp(BX) \cdot \operatorname{erfc}\left[\frac{X+((q/nR)^2 + (4\lambda D_\ell/R))^{0.5}t}{2(D_\ell t/R)^{0.5}}\right]\right] \qquad (9)$$

From Bear (1979) p 268

where: $B^2 = \left(\frac{q}{2nD_\ell}\right)^2 + \frac{\lambda R}{D_\ell}$

The organization of the model has been set up to utilize specified values for the hydraulic and chemical parameters of the flow system: bulk density, porosity, hydraulic gradient, hydraulic conductivity, distribution coefficient, dispersivity and radioactive decay constant. The parameters may have fixed values or a probability density function specified by mean, standard deviation, minimum, maximum and distribution type (i.e. Poisson, Uniform, Gaussian, Triangular, Beta).

For a particular trial using the model, single values of probabilistic parameters are selected at random by a Monte Carlo technique, and used to solve equation (9) for a specified time and distance. The distance refers to the point where groundwater from the tailings area discharges to a surface water body. In the case of this study, the reference tailings site has a groundwater flow path of 200 m.

The model has been adopted in Phase II to allow a time variant concentration source to be utilized. The conceptual basis applied in the groundwater model is that for a particular source function (result from the Source Term Model) a set of random parameters is used to determine the concentration of the contaminant at the discharge point. One trial run of the groundwater component model would correspond to the source function determined from a run of the Source Term Model. At this time, the groundwater component model requires a set of discrete data points comprised of concentration at a particular time, (i.e. t_i, $C(t_i)$) as input to the model. A set of data points representing a different source function would be required for each trial. For each source function, new parameter values are selected at random.

3. MODELLING TIME VARIANT INPUT

Contaminant transfer is described by a linear continuous partial differential equation. An increase in source concentration by some factor will cause a proportional increase in contaminant concentration at any distance and at any time. If a contaminant source can be represented as the sum of two (or more) sources, the resulting response can be represented as a sum of responses to each source. This is a consequence of the linearity of the contaminant transport equation and can be applied by using principles of superposition.

Any non-stationary (i.e. time variant) source, as in Figure 2a, can be approximated as a sum of elementary stationary sources originating at discrete time intervals. The numerical approximation of a time-variant source function is illustrated in Figure 2b and is summarized below.

$$\begin{aligned}
\underset{\substack{\text{Source at}\\ t_k}}{C(t_k)} &= \underset{\substack{\text{Source at}\\ t_o}}{C(t_o)} + \underset{\substack{\text{Source at}\\ t_1}}{[C(t_1) - C(t_o)]} \\
&\quad + \underset{\substack{\text{Source at}\\ t_2}}{[C(t_2) - C(t_1)]} + \underset{\substack{\text{Source at}\\ t_3}}{[C(t_3) - C(t_2)]} \\
&\quad + \cdots + \underset{\substack{\text{Source at}\\ t_k}}{[C(t_k) - C(t_{k-1})]} \qquad (10)
\end{aligned}$$

This type of approximation will have increasing accuracy as the incremental change decreases. The series of concentration changes can be thought of as elemental step functions which is equivalent to several time-constant sources of different concentrations.

Therefore, the solution for a time-variant source will be equivalent to the sum of the elementary solution for a stationary unit source, G(X,t), multiplied by the incremental concentration changes due to the non-stationary source as below:

$$C(X,t) = C(t_o)\ G(X,t) + [C(t_1) - C(t_o)]\ G(X,t-t_1)$$

$$+ [C(t_2) - C(t_1)]\ G(X,t-t_2) + \ldots$$

$$C(X,t) = C(t_o)\ G(X,t) + \sum_{0<t_k<t} [C(t_k) - C(t_{k-1})]\ G(X,t-t_k) \tag{11}$$

As discretization increases and increments tend to zero, the source function becomes equivalent to the integral:

$$C(X,t) = \frac{\partial}{\partial t} \int_0^t G(X,t-T)\ C(T)\ dT \tag{12}$$

where: T is the time increment.

Since the solution of the integral is difficult analytically, evaluation of the non-stationary source must be carried out numerically and the groundwater model calculates equation (11) rather than (12). An evaluation of the effects of discretization is therefore of importance and is discussed below in Section 4.3.

4. MODELLING DEMONSTRATION

In order to demonstrate the capability of the model to handle a time variant source term, several examples will be presented using theoretical ^{226}Radium source functions. These examples have been carried out in a deterministic fashion so that the results will not be clouded by calculations using randomly selected variables. All source functions used were defined as discrete step functions which reduces the error associated with approximating the area under a curve.

4.1 Simple Source Functions

Five simple cases were considered, and modelled over a 10,000 year period. Maximum input concentration used was 50 x 10^3 Becquerels/metre^3. The discharge concentration was calculated at various times using the parameters and values of Table 1 for a fixed distance of 200 metres. The breakthrough curves and associated source function are shown in Figures 3 to 7.

The five cases were:

(1) Constant Source (see Figure 3). For a constant source of concentration 50,000 Bq/m^3, the discharge concentration increases rapidly after 200 years reaching a maximum steady-state concentration (38720 Bq/m^3) by 2500 years.

The delay in arrival of the contaminant front is consistent with what would be expected for ^{226}Ra under the specified flow regime (see Table 1). A distribution coefficient of 25 corresponds to a retardation factor of about 161, which results in the arrival of the front ($C/C_o = 0.5$) after about 610 years. The steady-state concentration at the discharge is also consistent with the value expected due to radioactive decay after a 610 year period.

(2) Decreasing Pulse (see Figure 4). This example demonstrates the effects of a sudden decrease in source concentration. The breakthrough curve is consistent with this showing the arrival of the initial pulse, as in Case 1, followed by a rapid decrease to a steady-state concentration.

(3) Increasing Pulse (see Figure 5). This example demonstrates the effects of a sudden increase in source concentration. The two separate fronts can be seen clearly and are offset correctly.

(4) Increasing-Decreasing Step Pulses (see Figure 6).

(5) Decreasing-Increasing Step Pulses (see Figure 7).

Cases 4 and 5 both demonstrate the ability of the model to handle multiple concentration changes which may increase or decrease. The breakthrough curves lack some detail due to the number of points used, however, the delay in arrival and smoothing of the fronts is as expected.

4.2 Hypothetical Source Models

Two hypothetical source models have been developed by SENES which may be used to represent the concentration of Radium-226 (in Bq/m^3) input to the groundwater regime beneath a tailings pile. The two models are shown in Figures 8 and 9.

Model 1 is a linear model characterized by two distinct relationships:

for $0 \leq t \leq 200$ years
$C(t) = 5{,}000 + 225t$

for $200 \leq t \leq 10{,}000$ years
$C(t) = 50{,}000 - 5(t-200)$

The source strength increases rapidly to a peak, then decreases linearly through time.

Model 2 is perhaps a closer representation of the input source term and is a mixed model of linear and exponential components:

for $0 \leq t \leq 200$ years
$C(t) = 5{,}000 + 1.125t^2$

for $200 \leq t \leq 5{,}000$ years
$C(t) = 50{,}000 - (1.085 \times 10^{-3})\ (t-200)^2$

for $5{,}000 \leq t \leq 10{,}000$ years
$C(t) = 25{,}000 \exp[-6.458 \times 10^{-4}\ (t-5{,}000)]$

The three components represent three different changes in the function's slope which must be handled by the model as the concentration reaches a peak and then declines. The two models were used to prepare input data files for the groundwater component model. Several sets of data with different time increments were used to compare the effect of discretization on the discharge concentration (see Section 4.3).

The breakthrough curves for each of these source models are shown on Figure 10. These curves were calculated for a flow path length of 200 metres from the tailings and parameter values listed in Table 1. The curves are very similar in shape to the source models but diminished in concentration due to radioactive decay and smoothed by dispersive processes. The time increment used to define the source functions was 100 years.

4.3 The Effect of Discretization

As with all modelling exercises, there must be a certain "trade-off" between the accuracy obtained by utilizing a small time increment and the resulting economic, memory storage or calculation time constraints associated with the increment size. A high level of discretization will yield a more accurate result as the time increment approaches zero.

Several time increments were used to calculate the concentration at the discharge point at a certain time for each source model (X = 200 m, t = 10000 y). The results are shown in Figure 11a. It can be seen clearly that the resulting concentrations converge as the time increment decreases, although at a different rate for each source model. Computer storage restrictions did not permit the use of a time increment less than 11 years.

If it is assumed that the result using the smallest time increment (11 years) represents an accurate solution, then the other results can be compared to that solution. Plotting of this "% Error" against time increment as on Figure 11b shows that the numerical approximation and calculation approach yields results within less than 10 per cent of each other for time increments up to 100 years. There is some evidence that a numerical approximation of a time-variant source function is better suited to some curve shapes than others. For example, an exponentially declining curve will have less area between the curve and the incremental step function (i.e. less error) than a linearly declining curve. Thus, we see less discretization effects for the Model 2 function than Model 1 (see Figure 10). However, for either case it is recommended that a maximum time increment of 100 years be used for future modelling runs.

5. REVIEW OF PARAMETERS AND DISTRIBUTIONS

The specification of values for input parameters is no less important for probabilistic assessments than deterministic approaches. An evaluation of the input parameter values should not only include probability density functions, where appropriate, reflecting uncertainties in data due to inhomogeneities or measurement but also possible functional correlations between parameters. Input parameter distributions required for the groundwater model are shown in Figure 1. It must be noted that most parameters are highly dependent on the geologic properties of the setting. Only the radioactive decay constant can be considered to be a fixed independent parameter.

The parameters are reviewed below in the context of a reference site located in a drainage basin area on the Canadian Precambrian Shield.

(1) Hydraulic Conductivity. There is a large body of direct and indirect evidence supporting a log normal probability density function for hydraulic conductivity (summarized by Freeze, 1975). This has been observed in a variety of materials from unconsolidated soils to rock cores. It is also well known that grain size distributions are log normal and that hydraulic conductivity is related to the intrinsic permeability of the porous media and hence, the grain size.

Hydraulic conductivity (K) is a function of not only the porous media but also the fluid permeating the media.

$$K = \frac{Cd^2 \rho g}{\mu} \qquad (13)$$

where: ρ is fluid density
μ is fluid viscosity
g is gravitational constant
d is mean grain diameter
C is a proportionality constant

For the glaciofluvial sandy aquifers often found in the Elliot Lake area, estimation and measurement of hydraulic conductivity are in the range of 10^{-4} m/s (3150 m/a) to 10^{-5} m/s (315 m/a) (Blair et al., 1980, Morin et al., 1982). A suitable mean value would be 5 x 10^{-5} m/s (1580 m/a).

If K is log normally distributed, then Y = log K will be normally distributed. Freeze (1975) has tabulated frequency distributions for many thousand hydraulic conductivity values and standard deviations for Y have an average value of 0.67 for sands and gravels. Therefore, a suitable probability distribution for hydraulic conductivity should be expressed as Y = log K, having a normal distribution with mean 3.20 (equivalent to 1580 m/a) and standard deviation of 0.67.

Smith and Schwartz (1981) have pointed out the critical role that spatial variations in hydraulic conductivity have in controlling contaminant transport. The groundwater component model does not consider these inhomogeneities. The model assumes a uniform, homogeneous porous media with a value for hydraulic conductivity randomly selected from the probability distribution function for each trial.

(2) Hydraulic Gradient. The hydraulic gradients (I) in sandy aquifers are often in the order of 0.01 to 0.005 but will be site-specific and likely non-linear. For the

purposes of this model a linear gradient is assumed. Uncertainty in hydraulic gradient data is often due to measurement or extrapolation errors between data points. Such errors can be considered to be normally distributed.

A high gradient of 0.05 has been reported by Morin et al. (1982) and will be taken as an average conservative value. Errors are unlikely to be greater than 20 per cent, so an arbitrary standard deviation of 0.01 would be reasonable.

For hydrogeologic flow regimes where the groundwater flux is constant, the hydraulic gradient can be correlated to hydraulic conductivity using the relationship Q = KIA, where Q is the discharge rate, and A is the cross-sectional area through which flow occurs. Gradient in this case will be inversely proportional to the hydraulic conductivity. However, the groundwater component model assumes a constant velocity and flux for each particular trial calculated from the randomly selected parameters, hydraulic conductivity, hydraulic gradient and porosity.

(3) Porosity. The porosity (n) of a porous medium is the ratio of void volume to total unit volume. Porosity of unconsolidated materials will be dependent on the texture and grain size distribution of the deposit. In general, porosity ranges from 25 - 50 per cent, with coarse rounded materials having lower porosities than platy clayey materials. Very little data are available concerning probability distributions. However, it would be reasonable to assume that there is some relationship with grain size which is log normally distributed. Sandy aquifers commonly have porosities between 0.30 and 0.35. A representative mean value for modelling purposes would therefore be 0.33.

For a log normal distribution where Y = log n, these values correspond to a mean Y value of -0.48 with a standard deviation of 0.04.

In general, higher porosity materials (e.g. clay) have lower hydraulic conductivity.

(4) Bulk Density. The bulk mass density (ρb) of a material is the ratio of the total mass to the total volume. Typical values of bulk density usually fall in the range of 2.1 to 1.5 for most mineral soils. The bulk density is related to the specific gravity and porosity of the media by the relation:

$$\rho b = \rho s\ (1 - n) \tag{14}$$

where ρs is the particle mass density, which can be considered equivalent to specific gravity. This allows a correlation between porosity and bulk density where bulk density can be considered dependent on the porosity, i.e. an increase in porosity must cause a decrease in bulk density. Assuming that porosity is linearly related to grain size, which is log normally distributed, the probability density function for bulk density can also be assumed to be log normal.

In most cases it can be assumed that $\rho s = 2.65\ g/cm^3$. Morin et al. (1982) have reported that a sandy aquifer near a tailings area had a significantly higher specific gravity for the solid phase minerals of $2.98\ g/cm^3$. This value and the specified distribution for porosity give a mean ρb of 2.00 with a range from 1.9 to 2.1. For a log normal distribution where Y = log ρb, these values correspond to a mean Y value of 0.30 with a standard deviation of 0.02.

(5) Distribution coefficient. The distribution coefficient (Kd) is the ratio of the concentration adsorbed to the solid phase to the concentration in solution. The coefficient is very much dependent on soil type and pH, being more mobile at low pH. Values of Kd for ^{226}Ra have been reported in the literature covering the entire range from 0 to several thousand, however, Kd values are generally accepted to be in the range 25 to 250 mL/g. There is some uncertainty in the applicability of the batch test method of determining Kd. However, it is evident that ^{226}Ra will be strongly retarded in saturated unconsolidated deposits. In situ tests are not yet fully developed and few field observations have been reported. Morin et al. (1982) have calculated a Kd of about 70 mL/g based on observations from a contaminant plume in the Elliot Lake area. This appears to be a reasonable average value which fits into the range of laboratory estimates.

This Kd value was determined from observations within the inner acidified core of the tailings plume. In this zone, the mobility of radionuclides is controlled by the acid neutralization capacity of the aquifer. Morin et al. (1982) have also estimated that the migration of the neutralization zone is about 9 per cent of the groundwater velocity. Contaminants such as ^{226}Ra will be held back by the rise in pH at the acidic front by adsorption or coprecipitation to be later released as the acidic plume advances. Depending on the neutralization properties of the aquifer it may be expected that the arrival of an acidic plume discharging to a surface water body would be marked by an initial high concentration of ^{226}Ra which would rapidly decrease to a concentration representative of the source input and mobility controls of the acidic zone.

The modelling approach described in this paper is to use Kd values which can be considered to be representative of flow in an acidified plume, which is a conservative estimate. It is also significant to note that the migration time for a retarded acid front in a low buffering capacity aquifer, such as near Elliot Lake, would be small compared to the proposed 10,000 year modelling period.

Estimation of a probability function distribution is more difficult than an average Kd value. It is reasonable to assume that Kd can be described by a normal distribution with a standard deviation of 30 mL/g. For a mean of 70 mL/g the effective range of Kd values is therefore 40 to 100 mL/g.

(6) Dispersivity. Dispersivity has been considered to be a property of the porous medium that is used as a measure of the medium's capability to cause mechanical dispersion. Unfortunately, unique values of dispersivities are not easily defined for a porous medium because dispersivity is scale-dependent, increasing with distance as the transported mass encounters larger scales of heterogeneity. This scale-dependency is one of the problems in applying the classical advection-dispersion equation for predictive purposes. The groundwater component model uses the value of dispersivity as a bulk equivalent parameter to describe spreading in a heterogeneous porous medium. The probabilistic density function for dispersivity should help to account for randomness in velocity variations due to large scale heterogeneities.

Anderson (1979) has presented values of regional dispersivities which range in value from 12 to 200 m for alluvial sediments. It appears that dispersivity values of 30 to 61 m were reported for flow systems of 305 to 450 m in length. Scaling down of dispersivity for the 200 m flow path considered in this study would suggest a reasonable dispersivity value of approximately 25 m. The probability distribution will be considered to be normal with an arbitrary standard deviation of 10 m.

(7) Radioactive Decay Constant. The decay of ^{226}Ra is a constant uniform process. The half-life of ^{226}Ra is about 1620 years, so the decay constant is approximately 0.00043 a^{-1}. The decay constant should be specified as fixed.

The distribution of these input parameters is summarized in Table 2.

6. SUMMARY

A groundwater component model using probabilistic techniques to describe contaminant transport has been developed. This model assumes one-dimensional, steady-state flow conditions incorporating contaminant-specific retardation parameters. The key output of the model is contaminant concentration at a specified time and distance.

The component model has demonstrated the ability to handle constant source or time-variant source input using the principle of superposition for a variety of source functions, including linear or mixed model source functions. Solutions presented were carried out using a deterministic approach.

The effect of discretization was examined using variable time increments for the input source function. Concentration at the discharge point converged as time increment decreased. Error due to numerical approximation was minimized for time increments less than 100 years.

The input parameters required for the model have been reviewed in terms of probability density function and correlation to other parameters. Parameters controlling hydraulic flow have been found to be best described by normal or log normal distributions, although specific site conditions are extremely important for both groundwater movement and contaminant migration.

The selection of input parameter distributions has been based on a theoretical uranium tailings basin which contains an unconsolidated sandy aquifer. Few case studies are available on the different environments which may occur on the Canadian Shield. The groundwater component model has not been prepared for contaminant transport in fractured bedrock, which may occur in some tailings areas, and must be used in view of the limitations inherent in the formulation of the partial differential equation describing contaminant transport.

REFERENCES

Anderson, M.P., 1979. Using Models to Simulate the Movement of Contaminants through Groundwater Flow Systems. CRC Critical Reviews in Environmental Control, Vol. 9, No. 2.

Bear, J., 1979. Hydraulics of Groundwater. McGraw-Hill Inc., Israel.

Blair, R.D., Cherry, J.A., Lim, T.P. and Vivyurka, A.J., 1980. Groundwater Monitoring and Contaminant Occurrence at an Abandoned Tailings Area, Elliot Lake, Ontario. Proceedings, 1st International Conference on Uranium Mine Waste Disposal, Vancouver. Society of Mining Engineers of the American Institute of Mining Engineers, New York, New York, pp 411-444.

Freeze, R.A., 1975. A Stochastic-Conceptual Analysis of One-Dimensional Groundwater Flow in Nonuniform Homogeneous Media. Water Res. Research, V.11, No. 5, pp 725-741.

Morin, K.A., Cherry, J.A., Lim, T.P. and Vivyurka, A.J., 1982. Contaminant Migration in a Sand Aquifer near an Inactive Uranium Tailings Impoundment, Elliot Lake, Ontario. Can. Geotech. Jnl., V.19, pp 49-62.

Smith, L. and Freeze, R.A., 1979. Stochastic Analysis of Steady-State Groundwater Flow in a Bounded Domain 2. Two-Dimensional Simulations. Water Res. Research, V.15, No. 6, pp 1543-1559.

Smith, L. and Schwartz, F.W., 1981. Mass Transport 3. Role of Hydraulic Conductivity Data in Prediction. Water Res. Research, V.17, No. 5, pp 1463-1479.

TABLE 1

INPUT PARAMETERS AND VALUES USED IN MODELLING EXERCISES

Parameter	Value	Units
Bulk Density (ρb)	1.92	g/mL
Porosity (n)	0.3	m^3/m^3
Hydraulic Gradient (I)	0.05	m/m
Hydraulic Conductivity (K)	315	m/a
Distribution Coefficient (Kd)	25	mL/g
Dispersivity (α)	25	m
Radioactive Decay Constant (λ)	0.00043	a^{-1}

TABLE 2

SUMMARY OF PARAMETER DISTRIBUTIONS

Parameter	Units	Distribution[3]	Notes
Hydraulic Conductivity (K) $Y = \log K$	m/a	Log-normal Y:N[3.20,0.671]	1
Hydraulic Gradient (I)	m/m	I:N[0.05,0.01]	1
Porosity (n) $Y = \log n$	m^3/m^3	log-normal Y:N[-0.48,0.04]	1
Bulk Density (ρb) $Y = \log \rho b$	g/mL	log-normal Y:N[0.30,0.02]	1
Distribution Coefficient (Kd)	mL/g	Kd:N[70,30]	1,2
Dispersivity (α)	m	α:N[25,10]	1
Radioactive Decay Constant (λ)	a^{-1}	Fixed $\lambda = 0.00043$	2

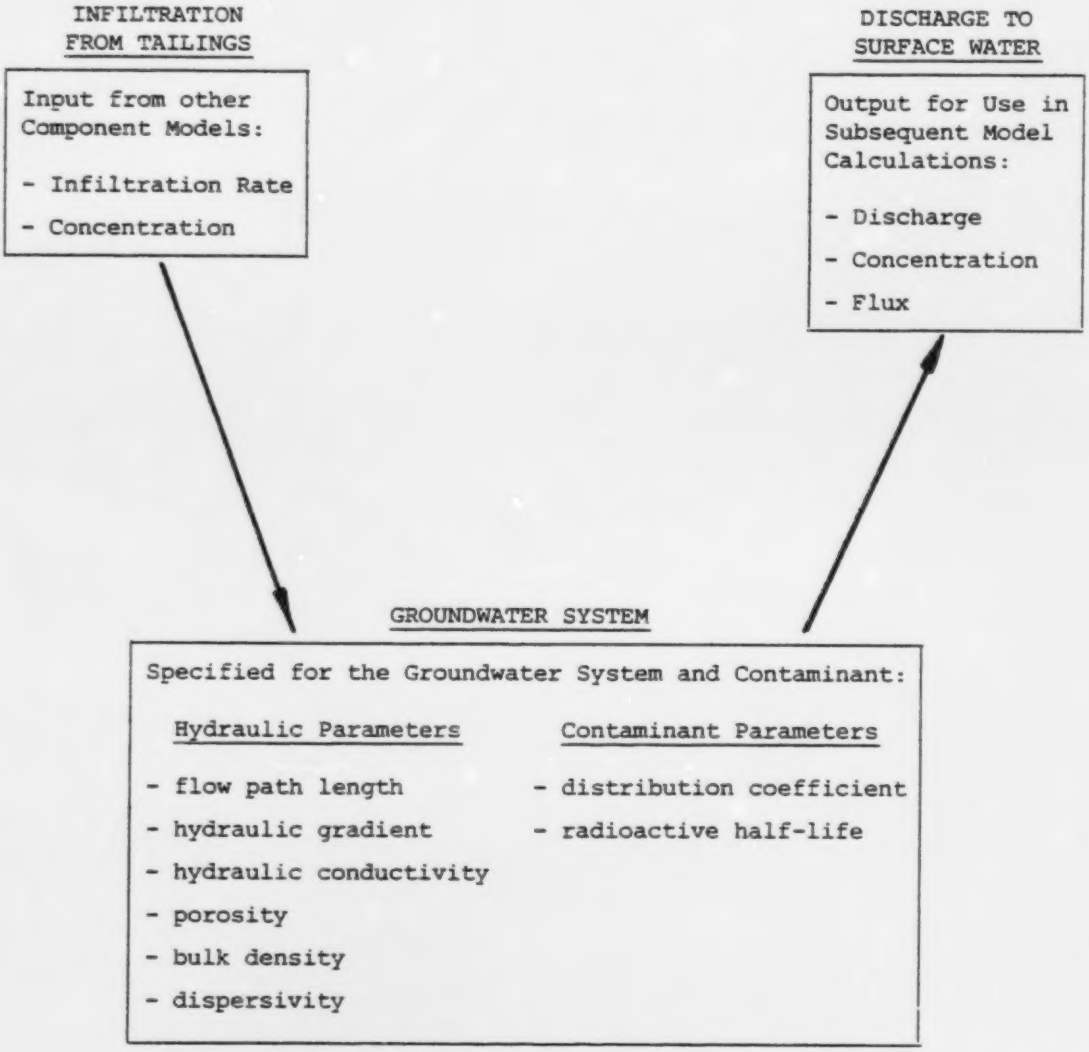
NOTES:

1) Dependent on setting

2) Dependent on contaminant

Distributions are presented in the form X:N[u,σ] meaning parameter X is normally distributed with mean u and standard deviation σ.

FIGURE 1: GROUNDWATER TRANSPORT COMPONENT MODEL CONFIGURATION, INPUT AND OUTPUT PARAMETERS



INFILTRATION FROM TAILINGS

Input from other Component Models:

- Infiltration Rate
- Concentration

DISCHARGE TO SURFACE WATER

Output for Use in Subsequent Model Calculations:

- Discharge
- Concentration
- Flux

GROUNDWATER SYSTEM

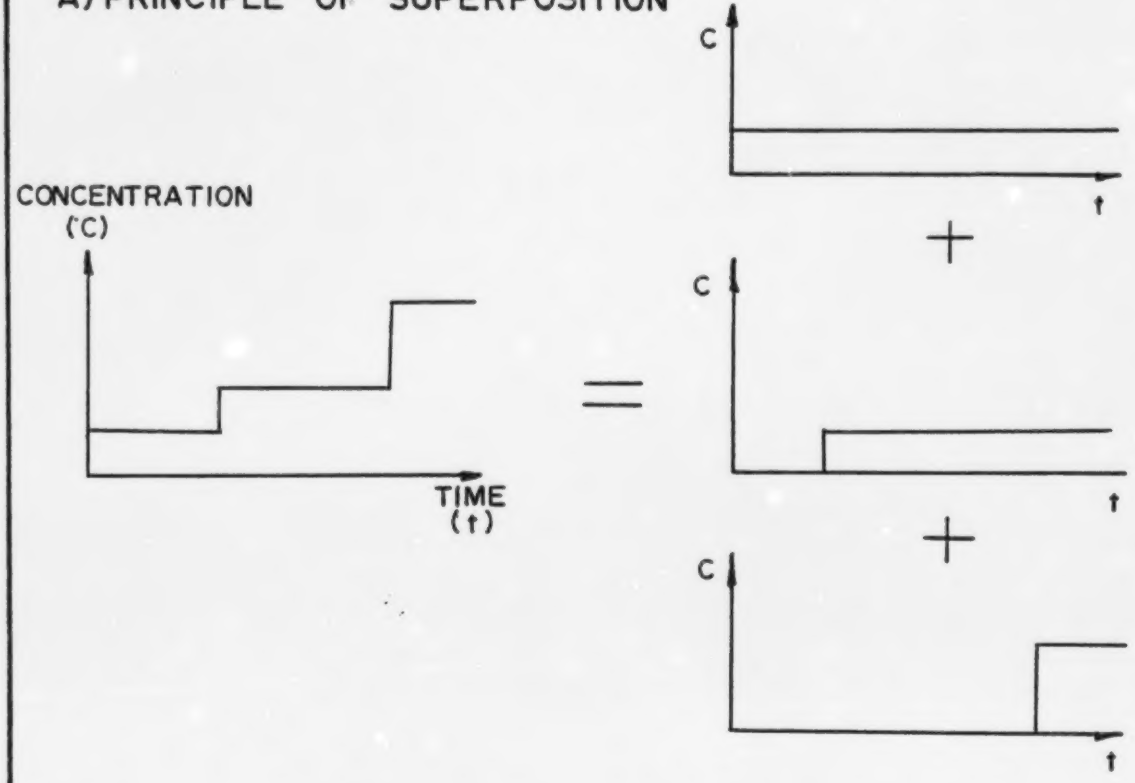
Specified for the Groundwater System and Contaminant:

Hydraulic Parameters	Contaminant Parameters
- flow path length	- distribution coefficient
- hydraulic gradient	- radioactive half-life
- hydraulic conductivity	
- porosity	
- bulk density	
- dispersivity	

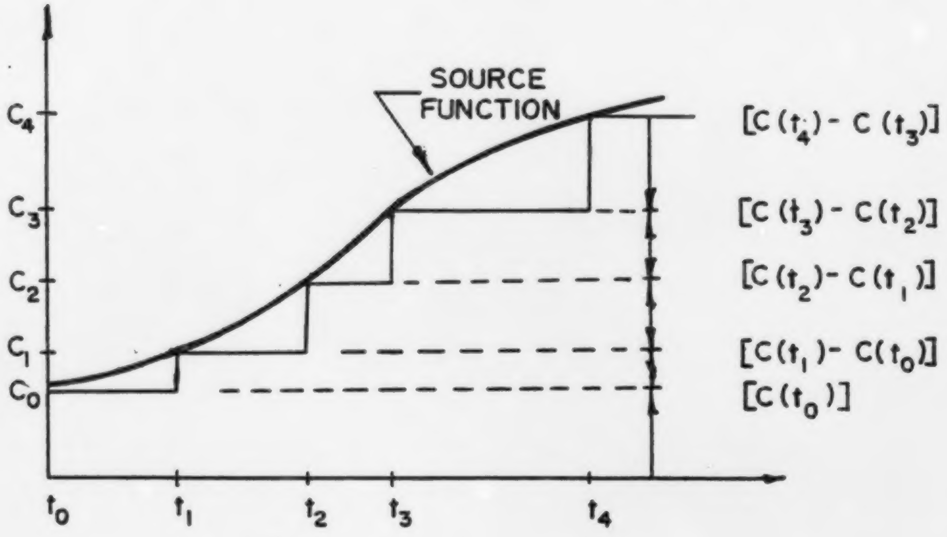
HANDLING TIME - VARIANT INPUT

FIGURE 2

A) PRINCIPLE OF SUPERPOSITION



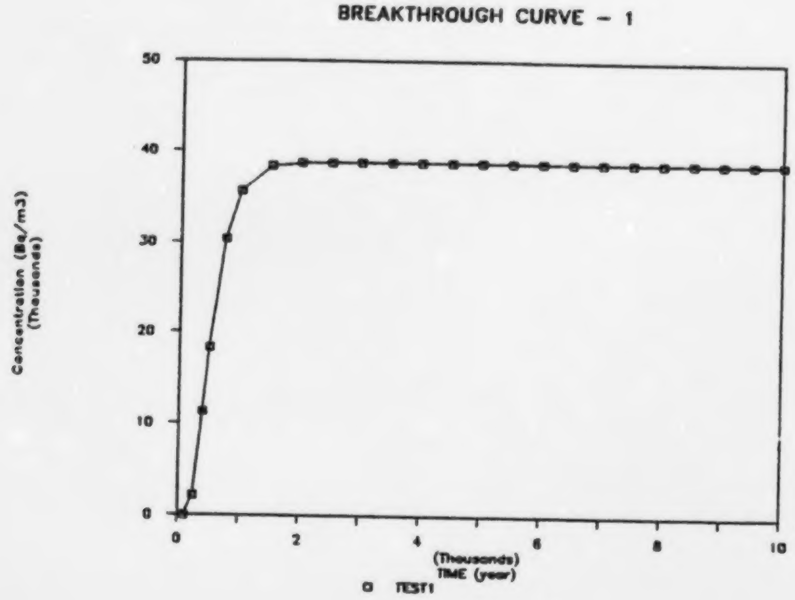
B) APPROXIMATION OF TIME - VARIANT SOURCE



Date JULY 15, 1985
Project No. 841-1328

Golder Associates

Drawn L.F.
Chkd.



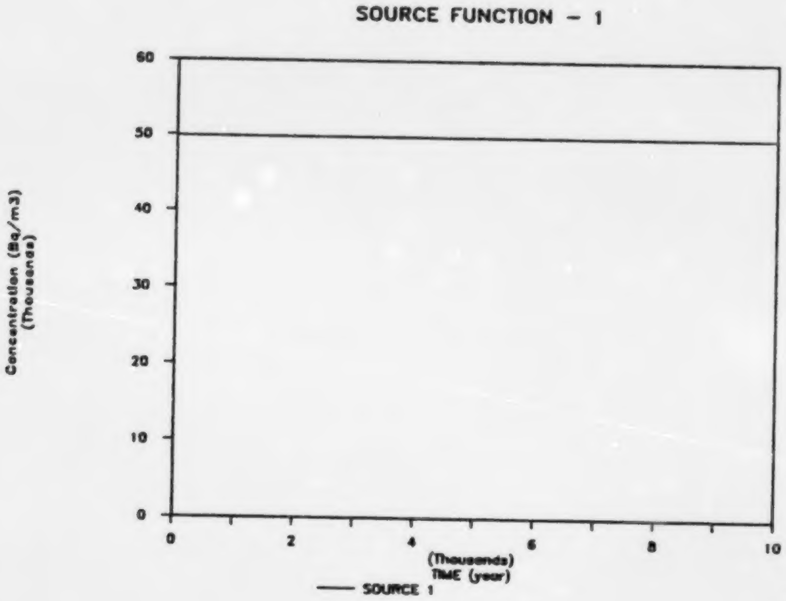
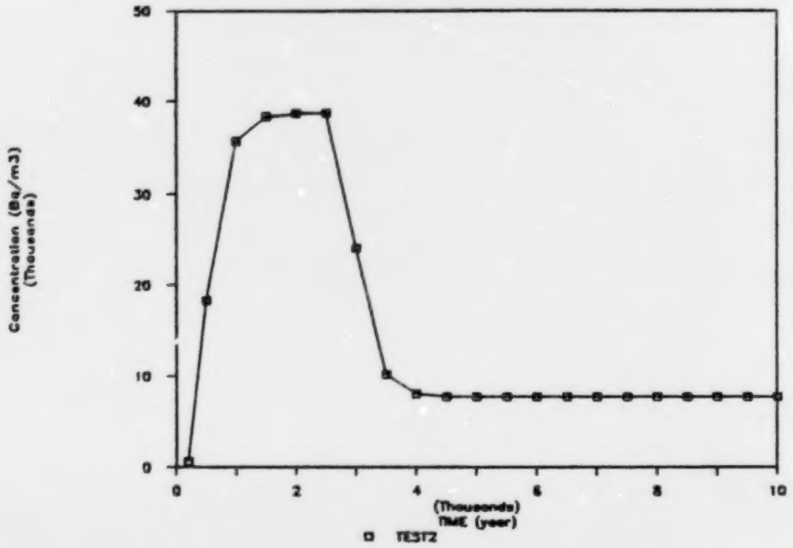
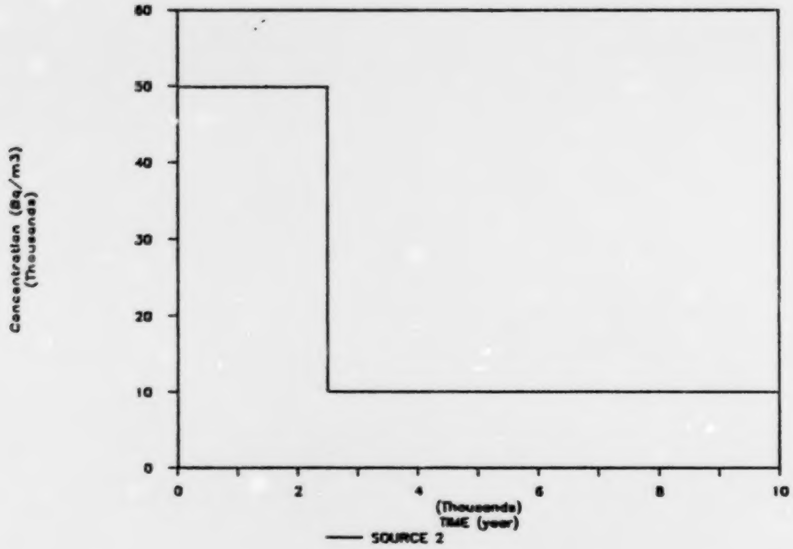


FIGURE 3: BREAKTHROUGH CURVE FOR A CONTINUOUS SOURCE



BREAKTHROUGH CURVE – 2



SOURCE FUNCTION – 2

FIGURE 4: BREAKTHROUGH CURVE FOR DECREASING PULSE

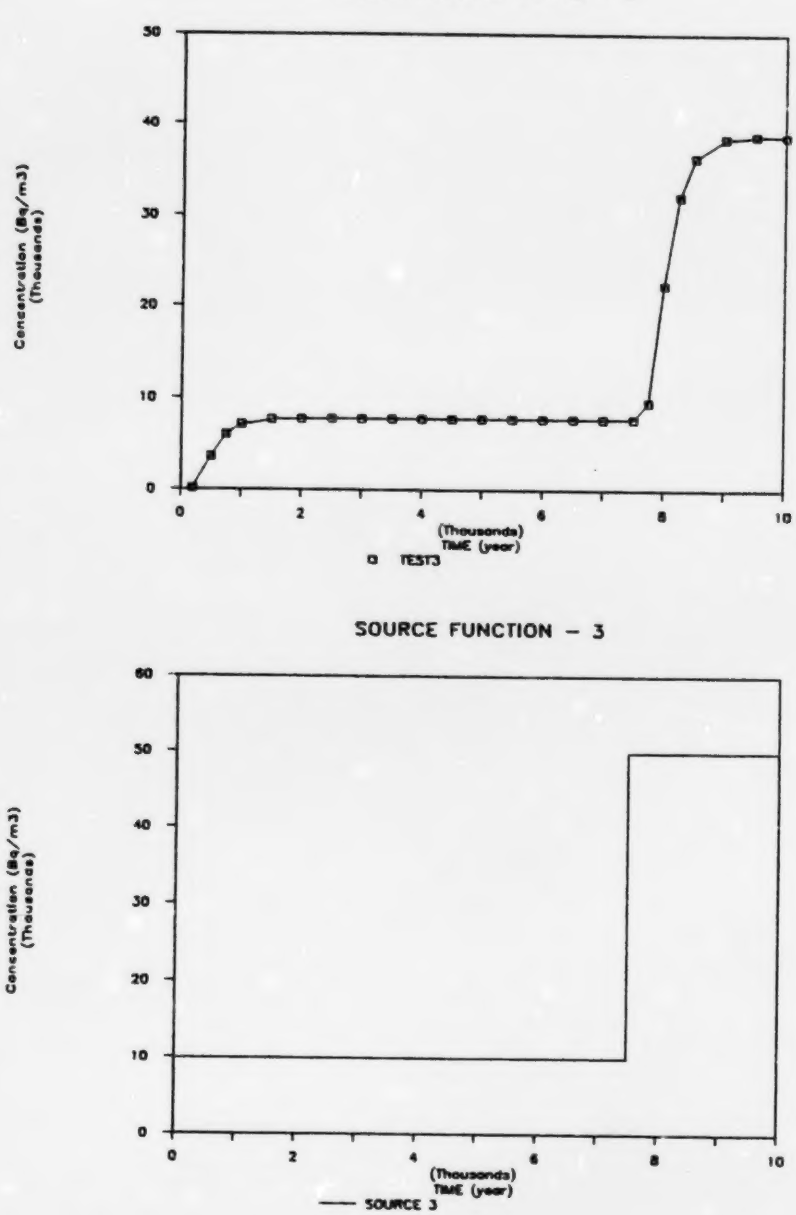
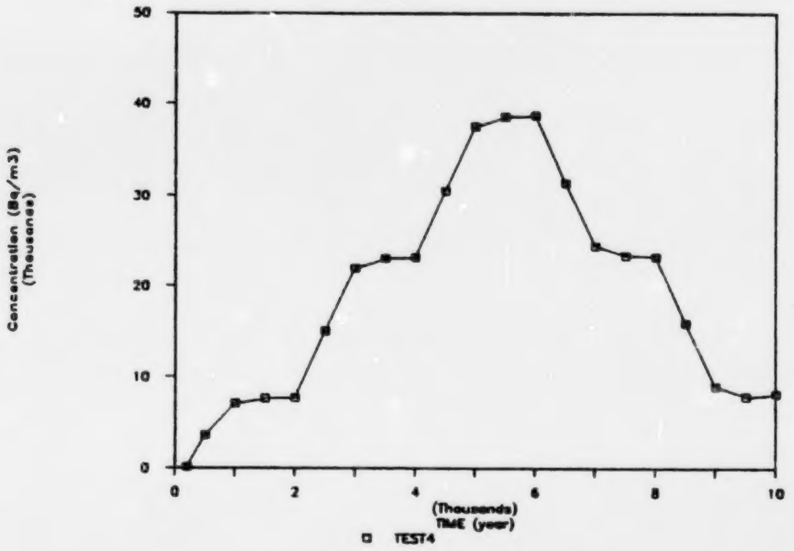


FIGURE 5: BREAKTHROUGH CURVE FOR INCREASING PULSE


BREAKTHROUGH CURVE – 4

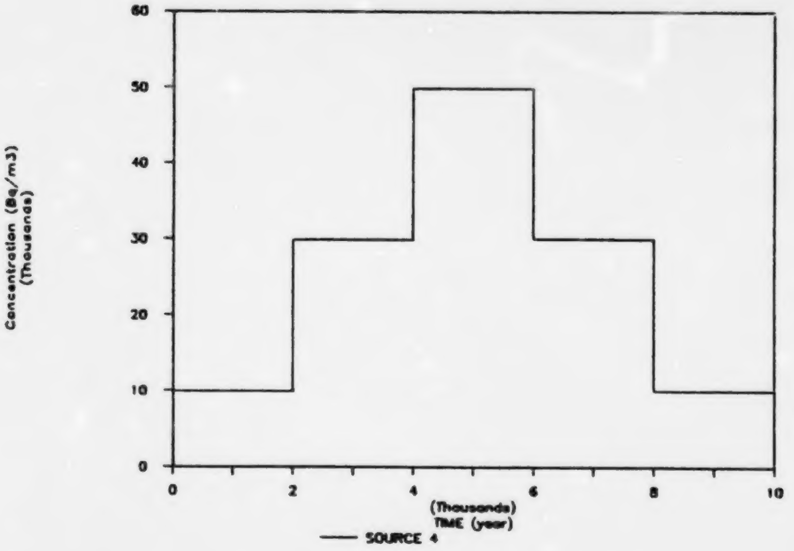

SOURCE FUNCTION – 4

FIGURE 6: BREAKTHROUGH CURVE FOR INCREASING/DECREASING PULSE

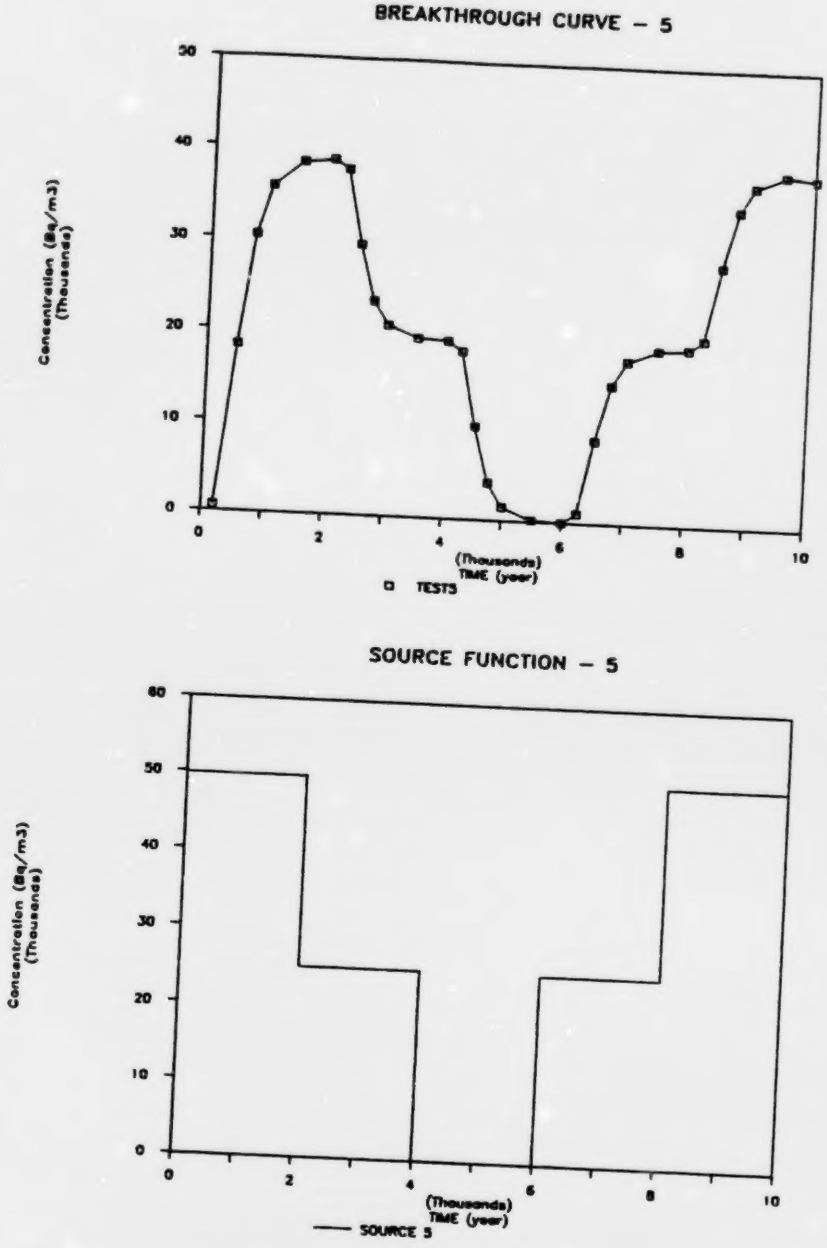


FIGURE 7: BREAKTHROUGH CURVE FOR DECREASING/INCREASING PULSE

FIGURE 8

LINEAR MODEL

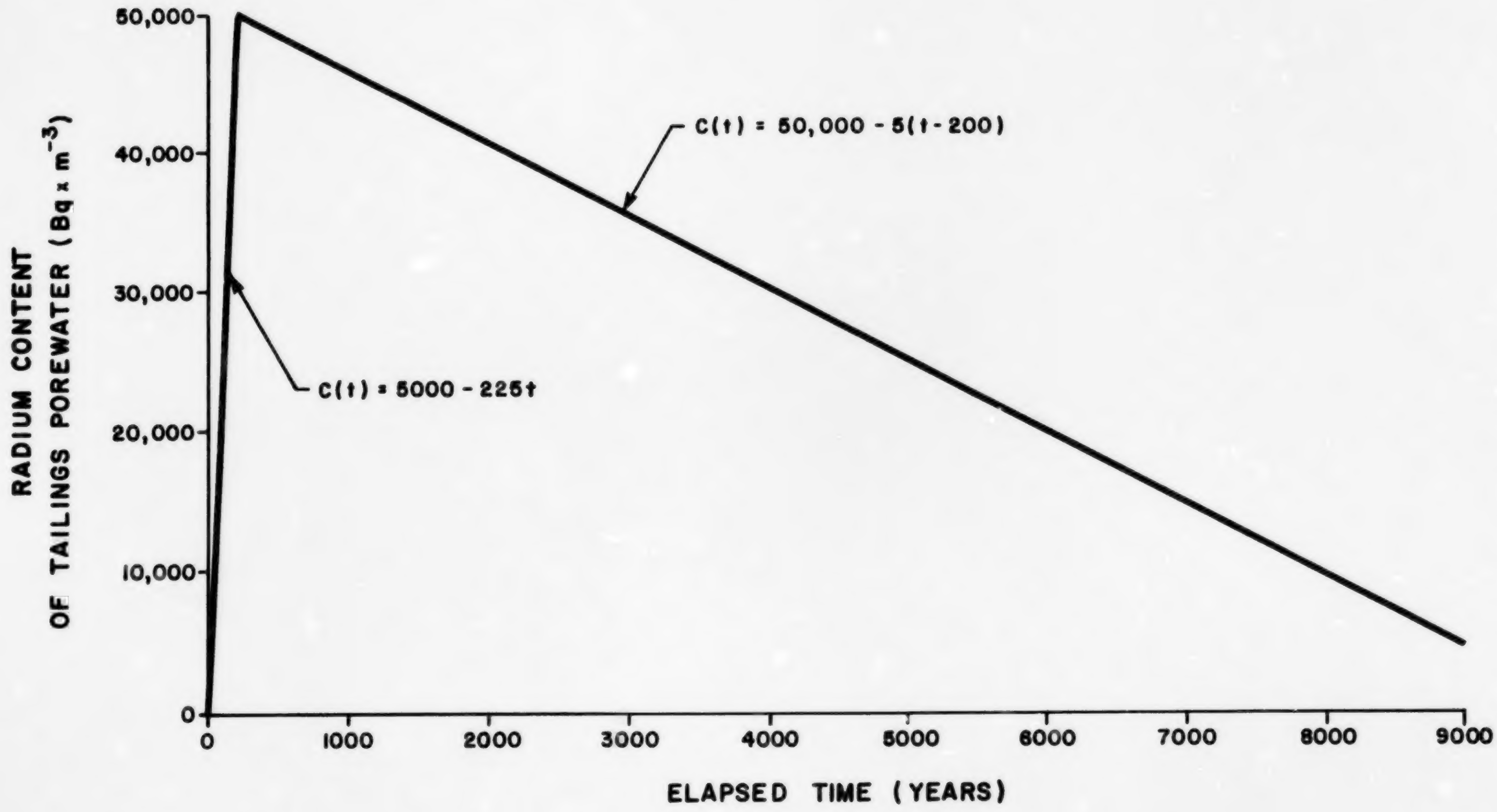
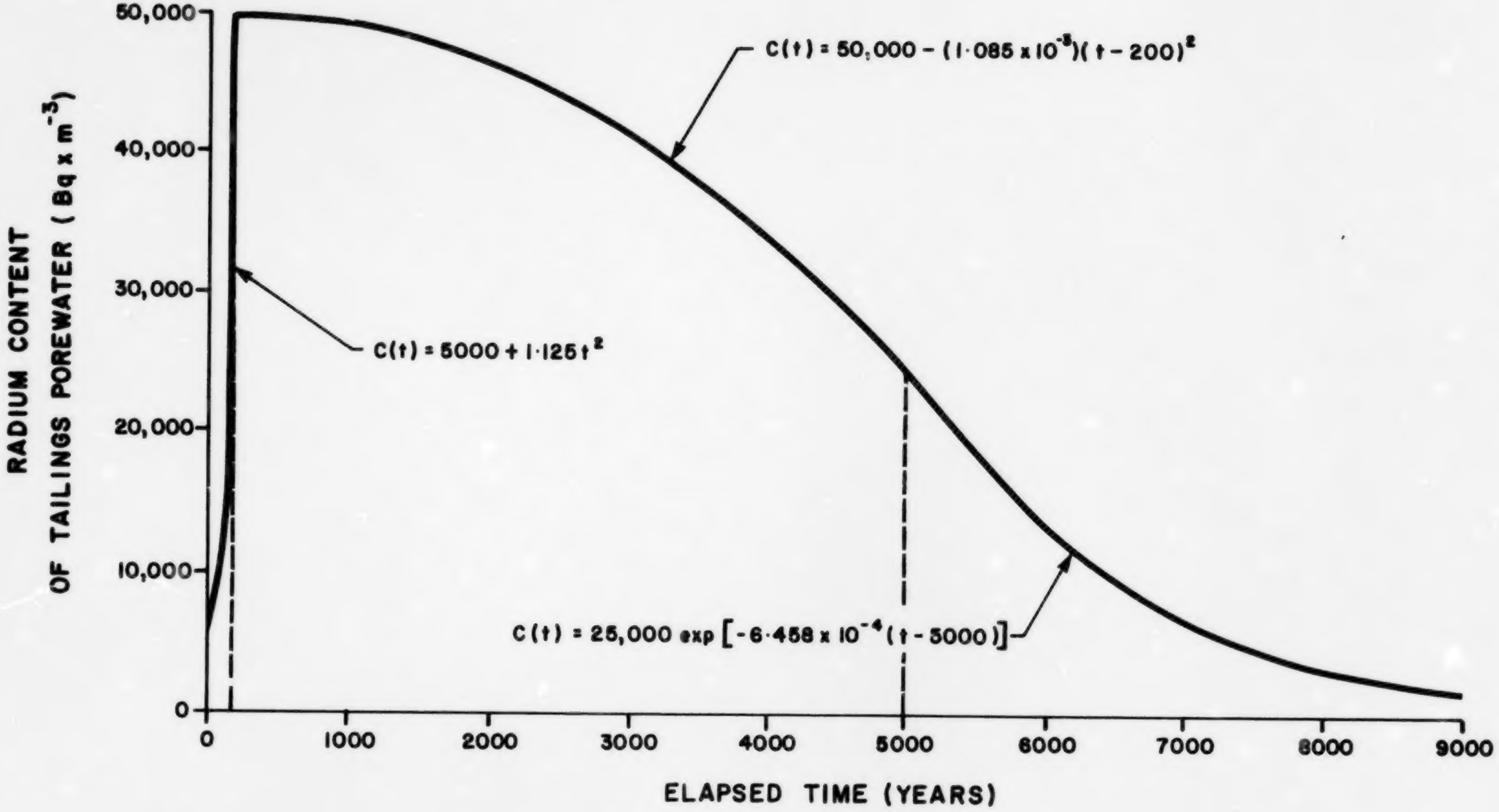


FIGURE 9

MIXED MODEL



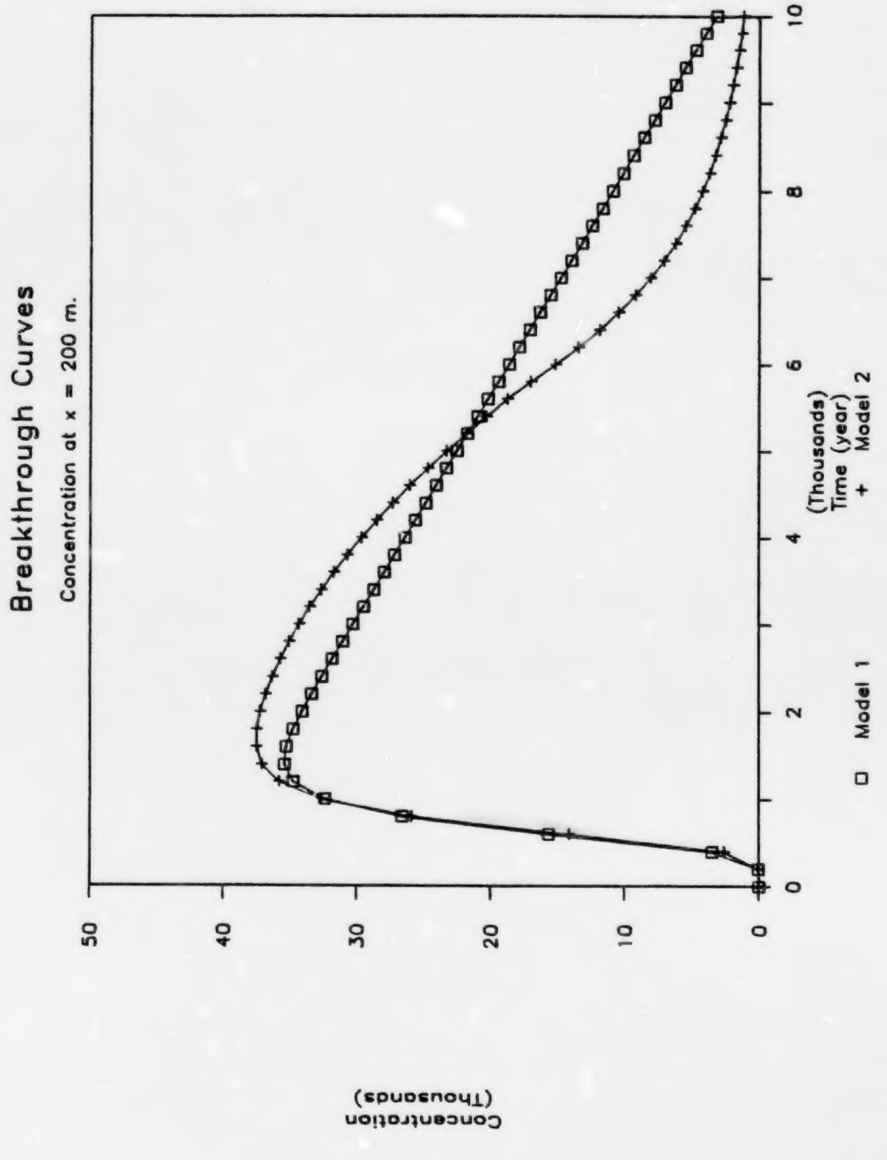


FIGURE 10: BREAKTHROUGH CURVES FOR HYPOTHETICAL SOURCE MODELS

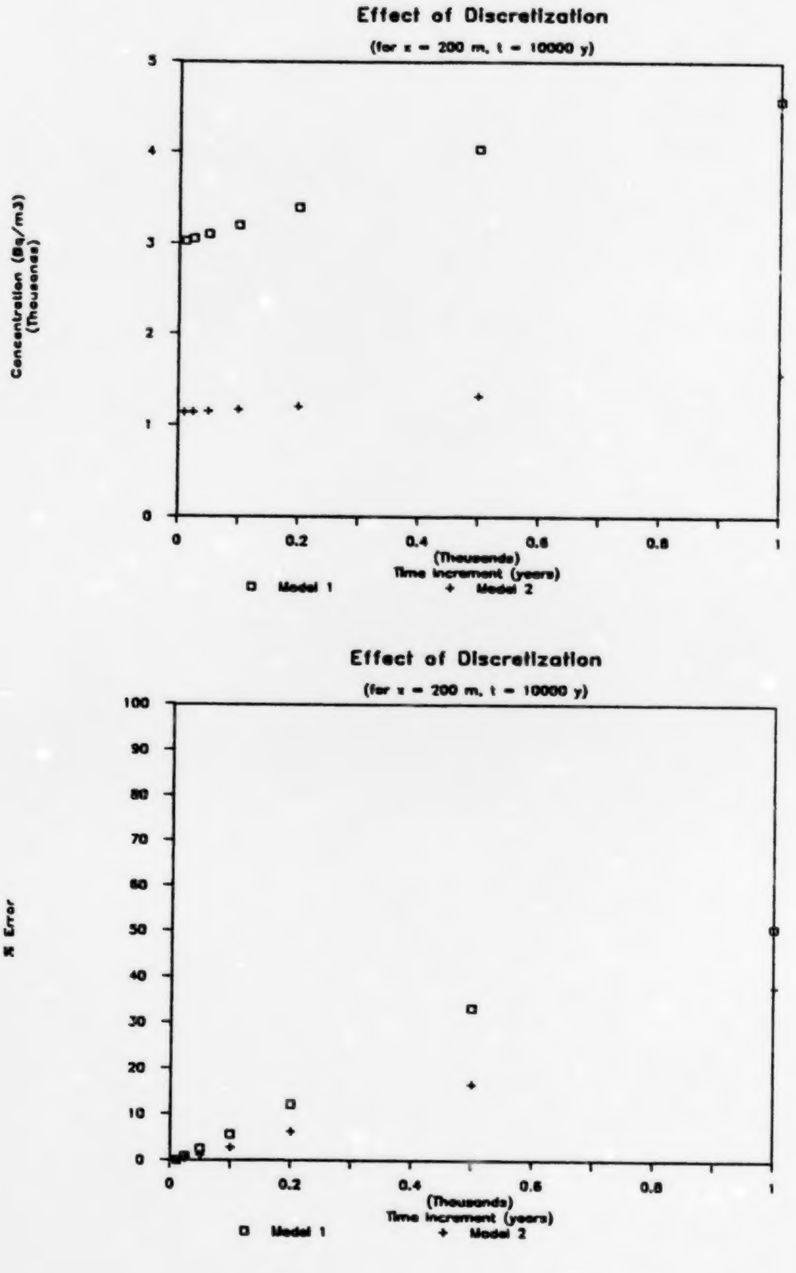


FIGURE 11: EFFECT OF DISCRETIZATION